An Evaluative Study of Some Selected Libraries in India Undergoing the Process of Digitization

Thesis submitted at the
Faculty of Arts, Jadavpur University, Kolkata
For the Award of Doctor of Philosophy

Anup Kumar Das

Jadavpur University
Kolkata, India
2008

Certified that the Thesis entitled

An Evaluative Study of Some Selected Libraries in India Undergoing the Process of Digitization submitted by me for the award of the Degree of Doctor of Philosophy in Arts at the Jadavpur University is based upon my own work carried out under the supervision of (1) Dr. Chaitali Dutta and (2) Dr. Bimal Kanti Sen and that neither this thesis nor any part of it has been submitted before for any degree or diploma anywhere/ elsewhere.

.....

(Dr. Chaitali Dutta)

Supervisor

.....

(Dr. Bimal Kanti Sen)

Supervisor

.....

Countersigned by the Supervisors

Dated: 11-09-2008

AK Das.

(Anup Kumar Das)

Candidate

Dated: 11/9/2008

To

My Beloved Parents

(Late Sri Ramesh Chandra Das and Late Smt. Bani Das)

Who Are Always in My Heart and Soul

Acknowledgement

This is high time for me to take this opportunity to express gratitude and thanks from the core of my heart to those without whose cooperation it would not at all possible to complete this task. At first, I want to express profuse thanks to my research supervisors Dr. Chaitali Dutta and Prof. Bimal Kanti Sen. They rendered expert guidance and generous help in the successful turnover of this work and assisted me in different ways by allowing exploiting different facilities on their own behalf.

I am highly obliged to my academic mentors Prof. S.B. Ghosh, Smt T. A. Pandalai and Late Smt Aruna Karanjai for their valuable discussions and suggestions about my topic. At many critical junctures in my life, the motherly affection of Smt Karanjai helped me in attainment of my professional goals.

I am extremely thankful to the respondents of my survey questionnaire, and key persons of the digitization initiatives across India who have been interviewed by me during my field study. They have also extended their friendly supports while demonstrating their digitization processes. In this connection, I would like to mention pioneering names in this nascent area who gave me a deep insight on the subject during my field visit at their institutions: Dr. Usha Mujoo Munshi of INSA, New Delhi; Dr. M.G. Sreekumar of IIM Kozhikode; Shri A. Hariharan of SERC, Chennai; Dr. Jagdish Arora of IIT Delhi; Prof. A.R.D. Prasad of DRTC, Bangalore; Dr. Vamshi Ambati of IIIT Hyderabad; Dr. V.N. Shukla of CDAC Noida; Prof. N. Balakrishnan of IISc Bangalore; Dr. S. Arunachalam of MSSRF, Chennai; and Dr. Ramesh C. Gaur of IGNCA, New Delhi.

My extreme gratitude finds no horizon to express her thankful acknowledgement to Ms. Jocelyne Josiah of UNESCO; Ms. Minja Yang of UNESCO; Prof. P.N. Desai, Chairperson, Centre for Studies in Science Policy (CSSP), Jawaharlal Nehru University (JNU); and other colleagues in JNU and UNESCO Office in New Delhi, who continuously encouraged me from different dimensions to complete this research work.

I am extremely thankful to Prof. Amitabha Chatterjee and Late Prof. Prabir Roy Chowdhury of Jadavpur University; Smt Kalpana Dasgupta of National Knowledge Commission for their valued guidance towards right orientation. The support from Dr. Krishnapada Majumder, Head, Department of Library and Information Science, Jadavpur University, at some critical junctures helped me to overcome some difficulties. I am taking this opportunity to express my gratitude to him. Necessary assistance from all other faculty and staff members of the Department of Library and Information Science, Jadavpur University, has been thankfully acknowledged.

Various academic discussions and useful tips from the following distinguished personalities are hereby gratefully acknowledged: Prof. Sujit Bhattacharya, Prof. V.V. Krishna, and Prof. S.K. Kejariwal of JNU; Prof. J.K. Sarkhel of Kalyani University; Dr. P.S. Mukhopadhyaya and Prof. S.C. Biswas of University of Burdwan; Shri Prakash Chand, Smt Nishy P. and Shri G. Mahesh of NISCAIR; Prof. Uma Kanjilal, Prof. Neena Talwar Kanungo and Dr. Jaideep Sharma of IGNOU; and Shri Ashok Sharma and Dr. Shankar Chowdhury of UNESCO.

I am expressing hearty thanks to my friends Bidyarthi, Bhaswati, Susmita-di, Indrani-di, Subir, Swapan, Sujit-da, Gopal-da, Aulie, Sanjay, Sabuj, Sevukan, Manidipa (Chotan), Dr. Premesh Sarkar, Dr. Jyotirmoy Samajder, Dr. Kaushik Sur, and other anonymous friends who stretched their friendly hands at some critical junctures to accelerate progress of this project.

This is the right time for me to remember my respected parents, my beloved sister (Late Smt Ranjana Das), my beloved Banima (Late Smt Bani Rani Ghosh, 1924-2002), who left me and was surely happiest persons at such buoyant moment of my life. I am also grateful to my uncle (Prof. Ranajit Kr. Dutta), parents-in-law (Shri S.N. Mukherjee and Smt Dipali Mukherjee) and brothers (Ashim and Ujjwal) for their incessant inspiration and cooperation in executing this job.

Last but not the least, my wife Smt Anasua sacrificed a lot for successful turnover of my applied effort and lent her hands at every moment.

Preface

Digitization of library materials provides an excellent opportunity to widely disseminate our documentary heritages and greatly increases access to library collections of rare documents as well as current research literature. Indian digitization initiatives aim at producing a vast amount of digitized documents pertaining to different forms of recorded human knowledge, ranging from the rare manuscripts to current research literature.

This work contains six chapters in all. The first chapter is 'Introduction', and it is sub-divided in different sections. The first section describes the background of the work. The next few sections have introduced with some fundamental concepts, for instance, open access, digital library, information retrieval, etc. These concepts form the basis of this study. In this chapter, some national and international policy instruments and guidelines along with the published literature by Indian practitioners are critically reviewed for better theoretical understandings of the subject.

Digital Library of India (DLI) is the largest digitization initiative in India spreading across states of India and involving over ninety organizations to ensure several thousands of rare books written in Indian languages as well as non-Indian languages are accessible through Internet channel. Chapter 2 critically appraises different aspects of the DLI project such as collaboration pattern, digitization processes, metadata practices, information retrieval, etc.

The higher education system in a country produces advanced research literature in the form of theses and dissertations in both thrust areas of the country as well as in the emerging academic fields. The researchers in the country need to consult results of past researches in order to expand frontiers of knowledge. Theses and dissertations are constant in demand for our academic research, although availability of the same is the matter of concern. Chapter 3 critically appraises different aspects of digitization work for theses and dissertations in higher learning institutions in India, more particularly initiatives such as Vidyanidhi Digital Library project, ETD@IISc and OpenMed@NIC.

Documentary heritage collections in the memory institutions in the country are on the verge of extinction. Chapter 4 critically appraises digitization programmes of two national level institutions covering the documentary heritage collections across states of India. While Indira Gandhi National Centre for the Arts (IGNCA) is engaged in multimedia documentation and digitization of multilingual documents in the areas of Indian culture, art and heritage, National Mission for Manuscripts focuses on most valuable collections of rare manuscripts covering many subject areas.

Journal literature is an important kind of research communication, where scholars publish their research findings. Indian journals can be found in almost every subject field – making a value addition to the world research literature. However, Indian journals were not available to the international research communities due to the weaker distribution channel. Indian journals as well as research papers published by Indian scholars in these journals were less visible to the international research communities. However, the research communication process has been drastically changed with the adoption of modern ICT-enabled publishing environment by the Indian journal publishers. Most of the publishers of Indian research journals belong to non-profit making scientific societies or institutions. Thus, open access to Indian journal literature was one of the viable options for making Indian journals internationally visible and available. Chapter 5 critically appraises digitization initiatives of some scientific institutions in India to make the retrospective volumes of Indian journals available. This chapter covers four major Indian open access journal initiatives, namely, e-Journals@INSA, Indian Academy of Sciences published journals, medIND@NIC, Medknow Publications.

Chapter 6, derived from the present research findings, concludes with a proposal of a theoretical framework for an inclusive national digital information system integrating all kinds of stakeholders, policy instruments, best practices and indicators for harnessing a knowledge society.

Table of Contents

List of Tables

List of Figures

Chapter 4

Chapter 5

CHAPTER 1

Introduction

Digital Library is where the past meets the present and creates the future...

Dr. A.P.J. Abdul Kalam

Former President of India

1.1 Background

India is one of the emerging economies having consistent economic growth rate in last two decades due to increasing contributions from the knowledge-based service sector. Leveraging prospect of liberalization and globalization of Indian economy that was introduced during early 1990s, India has become one of the top software exporting countries. While India exports cutting edge IT products and services to the developed nations, capabilities of other knowledge-based industries are also getting strengthened in order to meet global demand of knowledge-based services as well as stay ahead in the global competition (NKC, 2006). Knowledge-based industries are human capital intensive and strengthening knowledge networks across the country, and will definitely contribute to the development of human capital (NKC, 2007). However, knowledge economy in this nascent stage could only affect a segment of Indian society, while other segment is struggling for making a decent livelihood. The use of information and communication technologies (ICT) in the cross sections of society, and more particularly in areas such as human communication, knowledge dissemination, teaching, lifelong learning and governance is also getting strengthened to match the changing economic scenario.

During the post-liberalization period in India, a significant landscape changing cum nation building exercise has been undertaken with the establishment of a huge number of new professional and technical institutions and modernization of existing educational and research institutions across the Indian states in order to produce required human capitals essential for knowledge-based industries in

India. In this process, Indian institutions have strengthened their ICT infrastructure and ICT-enabled interactive learning environment.

In contrary to India's economic progress and prosperity, institutions providing public access to information and documentary resources to cross sections of people are on the other side of digital divide. Cultural and memory institutions in the rural, semi-urban and even urban areas across India are struggling to get operational ICT facilities due to several constraints such as lack of electricity, Internet connectivity, localized contents and trained manpower.

On the other hand, in South Asia particularly in India, large volumes of cultural heritage resources (documentary) are on the verge of extinction due to lack of preservation, non-availability, rarity and natural decay. The knowledge and wisdom lost from these cultural heritage documentary resources can cause a severe vacuum in the intellectual wealth and humanity, if we cannot preserve them at this critical juncture. The current information is also very much needed for the continuous development of the society and smooth functioning of our modern life. The information and communication technologies (ICT) act as a development enabler and intervener in uplifting quality of life through modernizing various functionalities of our life. The digital library is a product of recent innovations in the areas of ICT, where the information users can get full-text access to needed information at their workstations without delay, but with much satisfaction. Searching information from the physical documents is rather painful and time consuming if we compare it with electronic documents. The developed nations have already built up a number of digital libraries on various areas of studies. With the emergence of digital libraries, institutions across the world have strengthened their capability of serving users' communities with adequate information resources. In India, development of digital libraries has achieved some momentum in the late 1990s and early years of the twenty-first century. The national level institutions in India took keen interest in developing institutional repositories that are now available in the cyberspace, reaching corners of the world. Later, other institutions and organizations have also started initiatives at their institutional or organizational level to disseminate new

knowledge generated within the institutions or organizations. Availability of open source software has also accelerated these initiatives in India, which has resulted in another way of disseminating scholarly literature, i.e. open access literature. In contrast to subscription-based literature, the open access literature does not have any restriction on access, and is free from any subscription fee or licensing fee (Ghosh & Das, 2007).

India is ahead of many developing countries and a few developed countries in terms of establishing a number of digital libraries or digital archives and creating digital contents for them. World communities have appraised Indian efforts, and contents of some digital libraries are regularly accessed in different parts of the world (Ghosh & Das, 2007). India also plays a crucial role in the South Asia sub-region for the development of digital libraries, which reflects the cultural diversity of the people of south Asian countries and the traditional knowledge of this region (Das, 2008).

1.2 Concepts of Digitization and Digital Library

As defined by Reitz (2008), digitization is “the process of converting data to digital format for processing by a computer. In information systems, digitization usually refers to the conversion of printed text or images (photographs, illustrations, maps, etc.) into binary signals using some kind of scanning device that enables the result to be displayed on a computer screen.” She also defines digital library as the “library in which a significant proportion of the resources are available in machine-readable format (as opposed to print or microform), accessible by means of computers”.

The digital libraries store, organize and disseminate digital contents. These contents are created either through digitization of existing printed materials and media documents, or through re-keying/re-composing of existing printed materials and media documents, or through creating new documents in digital formats. The first kind of documents is known as digitized documents, and the later kind of documents is known as born digital documents. In Indian digital

libraries both kinds of documents are available. The digitized documents are stored either in image formats or in text formats. If the original documents are available in European languages such as English, French, German and Spanish, the optical character recognition (OCR) software can automatically convert them into searchable digital text format, where qualitative OCR conversion rate is much higher. On the other hand, if the original documents are available in Indian languages such as Sanskrit, Hindi, Bengali, Oriya, Telugu and Tamil, the contents are made available either in image formats or re-keying the texts for the inclusion in the digital libraries. OCR software for Indian languages is still in the developmental or testing stage, where OCR conversion rate is much lower than acceptable rate. The full-text searching is possible in textual documents but this facility is absent in image documents.

1.2.1 Documents and Collections in Digital Library Systems

Digital library is the concept of information stored digitally and made accessible to users through digital systems and networks, but having no single location. It is, therefore, analogous to a library as a storehouse of information, but has a virtual existence in the digital spaces. Digital library is essentially a fully automated information system with all resources in digial form. Many views of digital libraries stem from what libraries currently do. Traditional libraries collect, organise, provide access to, and preserve objects in their collections. A library collection may include books, magazines, journals, theses, dissertations, manuscripts, audio-visuals, maps, etc. The flexibility of digital technology allows it to handle new kinds of object efficiently. Digital library collections can include things without direct physical analogs, such as algorithms or real time data feeds. They also may include digitized representations of what have traditionally appeared largely in museums and archives. With the rise of cost of paper publications and library storage, increasing use of computers, decreasing budgets, many libraries have to reduce their acquisition of books as well as their journal subscriptions. Documents in electronic form can become more easily available and widely used because the cost of digital storage and processing is going down.

Documents are the heart of digital libraries. Without documents there would be no digital libraries. In digital libraries, documents are not only what are stored in traditional libraries (e.g., books, journals, pictures and videos), but also include many works uncommon to those libraries, e.g., multilingual, multimedia, and structured documents (e.g., books broken into chapters, sections, subsections, figures with attached captions, colour graphics or images, attached or linked sound or video files, appendices, indexes, and 'front matter'); programs, algorithms, bulletin board archives, besides others. A document can have various representations depending on its intended use; for example, some applications require high-resolution images of documents with invisible watermarks for security purposes as well as low-resolution images for children to download from the Internet. Collections of digital library ranges from small, self-contained, and narrowly defined collections to ones spread across physical and logical spaces. One of the common requirements for a digital library is the ability to deal with distributed collections of information.

1.2.2 Evaluation of Digitization Work and Digital Library System

A digital library may be evaluated from a number of perspectives, such as collaboration pattern, system, access and usability, user interfaces, information retrieval, content and domain, services, cost and overall benefits and impact. An important issue under discussion across various communities is the set of metrics to be used for evaluating digital libraries. Selection of digital library metrics should be considered from both system-oriented and user-oriented viewpoints. From the system's perspective, we consider capacity (number of digital objects stored and number of users served simultaneously), content, transaction speed (speed of search response). From the user's perspective, we consider impacts of the system on the user (e.g., impact on patterns of association and attitudes about the digital libraries), effectiveness (relevance of the results; ability to produce a ranked list of results that are mostly relevant with best matches at the top), usability (e.g., ease of use, suitability to purpose, user's effort), interactions with the system, and user satisfaction.

In a general way, the constructs or elements for evaluation of digitization projects covered in this study are:

- Collaboration pattern for collection building;
- Collaboration pattern for resources mobilization and utilization;
- Selection of contents for digitization;
- Digitization workflow;
- Interpretation, representation and metadata;
- Access and distribution – open access versus campus-wide (closed) access;
- User interfaces – search and retrieval; and
- Integration, cooperation with other resources and libraries.

1.3 Problem Definition

As digital library system is an integrated information system, comprising metadata, multilingual texts, images, audio-visuals and multimedia data and also an efficient communication system, this system follows the system development life cycle. Beyond the post-implementation review of a digital library system, there are issues such as management of digital content, availability and accessibility of digital library services, impact on users, impact on the organization, legal constraints, retrieval efficiency, etc. need to be assessed and measured. The present study addresses the some of the necessary aspects affecting the online and digital library systems in India.

Digitization and open access projects in India serve a diverse group of users and applications, and a particular solution might not be relevant to all users and applications. Possibility of incorporating different solutions has been studied in this research work. Indian digital libraries have diverse services and products based on the mission and objectives of the host organizations. For example, proposed digital library of the National Mission for Manuscripts will provide full-text access to rare Indian manuscripts. Digital library projects of Indira Gandhi National Centre for the Arts (IGNCA) provides online access to multimedia and multilingual documents pertaining to Indian art, culture and heritage.

On the other hand, a number of institutions of higher learning in India including Indian Institute of Science (IISc), Indian Statistical Institute (ISI), Indian Institutes of Technology (IITs), Indian Institutes of Management (IIMs), National Institutes of Technology (NITs) and Indian universities are digitizing their doctoral dissertations, research reports, research papers and other technical documents for incorporating into their own institutional repositories and ETD (electronic theses and dissertations) repositories. Although few digital libraries are available in public domain as compared to the number of digitization projects undertaken across in India, Internet technologies and new communication technologies have widened its horizon to the wider communities. It is necessary to assess and measure their impact to the user communities, to review the technologies and their limitations, and to address strategic issues such as fulfillment of missions and objectives set by the digitization projects. The present study tries to incorporate some of these important aspects of the Indian digitization projects. In this research work, evaluation and assessment of existing digitization projects have been carried out on a case by case basis to critically examine technical, operational as well as managerial components of each of them.

1.4 Objectives of the Study

The main objectives of this research study are:

- To evaluate and assess the national-level digitization and digital library projects in India;
- To evaluate and assess the national-level projects in India aiming at open access to knowledge and scholarly literature;
- To evaluate and assess the digitization projects for different types of collections such as rare books, manuscripts, journal articles, documentary heritage, theses and dissertations;
- To study critically the collaboration patterns in digitization projects in India including technical and financial collaborations;
- To examine the digitization workflows in different digital library projects.

- To examine international visibility and usability of open access digital library systems in India;
- To examine the use of metadata in the digitization projects for describing digitized contents;
- To study critically the search and retrieval interfaces adopted in the digitization and open access projects in India;
- To study critically the content management mechanisms adopted in the digitization projects; and
- To suggest the effective model for dissemination of indigenous knowledge, documentary heritage and current scholarly literature.

1.5 Scope

This research work aims at producing *State of the Art Report* in the emerging area based on exploratory research method. The objective of exploratory research is to gather preliminary information that will help define problems and suggest hypotheses. Given its fundamental nature, exploratory research often does not require any hypothesis (Kotler & Armstrong, 2006). Thus, this research work was carried out without any hypothesis.

In this thesis, some well-known projects have been evaluated based on some parameters as indicated in Section 1.4 of this Chapter. Apart from some national level initiatives, some institutional cases are also covered in this thesis while these institutional initiatives are strongly representing national ethos on this emerging area.

1.6 Methodology

This research work has been carried out in different phases.

Phase I: In this phase the author was involved in review of literature with special focus on national and international policy instruments, digitization standards, digitization guidelines, best practices, country reports, institutional case studies,

and current research findings related to digital library development in the country and abroad.

Phase II: In this Phase the author prepared a structured questionnaire and tested this questionnaire in consultation with digital library practitioners in India. After successful testing, questionnaires were circulated to project coordinators of selected digitization initiatives across India. This questionnaire is provided the Annex I of this thesis. Later, the author undertook field visits to the important digitization project sites across India for collecting data, applying observation and interview methods. The basis for selecting the institutes for study and field work was (i) representative project sites of national level initiatives such as Digital Library of India's regional mega scanning centres, scanning centres and source libraries; Indian Institute of Science (IISc), Indira Gandhi National Centre for the Arts (IGNCA), National Mission for Manuscripts (NMM), National Informatics Centre (NIC) and many other important institutions across the country; (ii) prior appointments with the key persons in the national level initiatives – whoever responded in time.

In this field work, the author interviewed project coordinators, project supervisors and project staff – whoever were available – for recording strategic, operational and practical information related to digitization work. In this phase critical and exhaustive studies were carried out mostly following case study method. Case studies revealed detailed descriptions of digital library projects. Each component in a digitization project set up such as ICT infrastructure, digitization equipment, software used for content digitization, quality control and aggregation; information storage and retrieval system are critically observed and recorded. Observations in the field visits were properly documented as per scientific research method.

Phase III: This phase revolves around critical analysis of gathered data. In this phase, final report is prepared in the form of thesis incorporating analytical description of selected digitization projects, recommendations and suggesting future research directions.

In this thesis *APA Style* is followed for citing references and sources of information, which is based on fifth edition of the Publication Manual of the American Psychological Association (APA, 2001).

1.7 Review of Literature

1.7.1 International and National Policy Instruments

The policy instruments are intervention tools used to achieve national and institutional objectives of a given thrust area. The key policy instruments for the open access to knowledge and information are international declarations accepted in intergovernmental forums, recommendations of policy advisors to highest government authorities and government regulations. National level organizations and local institutions may use a variety of policy instruments to achieve their objectives. It is very fortunate that Indian digitization and open access initiatives have been given much importance on various provisions of these policy instruments and adopted appropriate measures in implementing them.

1.7.1.1 International Policy Instruments

a) WSIS Declaration of Principles and Plan of Action

The United Nations-backed World Summit on the Information Society (WSIS) strongly supported open access to information and knowledge (United Nations, 2005). This confirms that member countries of the United Nations will take appropriate strategic decisions to bring scholarly literature, produced from public-funded research initiatives or state-supported researchers, under the umbrella of open access. In this instrument, some specific principles and plan of action were suggested relevant to digital preservation of cultural contents, creation of local contents and open access to scholarly literature. An excerpt of relevant texts from both *Declaration* and *Plan of Action* is highlighted below.

WSIS Declaration of Principles[1]

"We strive to promote universal access with equal opportunities for all to scientific knowledge and the creation and dissemination of scientific and technical information, including open access *initiatives for scientific publishing."[B3. Access to information and knowledge]*

"The preservation of cultural heritage is a crucial component of identity and self–understanding of individuals that links a community to its past. The Information Society should harness and preserve cultural heritage for the future by all appropriate methods, including digitisation*." [B8.Cultural diversity and identity, linguistic diversity and local content]*

WSIS Plan of Action[2]

"Support efforts to develop and use ICTs for the preservation of natural and, cultural heritage, keeping it accessible as a living part of today's culture. This includes developing systems for ensuring continued access to archived digital information and multimedia content in digital repositories*, and support archives, cultural collections and libraries as the memory of humankind.*

Develop and implement policies that preserve, affirm, respect and promote diversity of cultural expression and indigenous knowledge and traditions through the creation of varied information content and the use of different methods, including the digitization *of the educational, scientific and cultural heritage.*

Support local content development, translation and adaptation, digital archives, and diverse forms of digital and traditional media by local authorities. These activities can also strengthen local and indigenous communities." [C8. Cultural diversity and identity, linguistic diversity and local content]

"Encourage initiatives to facilitate access, including free and affordable access to open access *journals and books, and* open archives *for scientific information." [C3. Access to information and knowledge]*

[1] World Summit on the Information Society (WSIS) - Declaration of Principles <www.itu.int/wsis/docs/geneva/official/dop.html>
[2] World Summit on the Information Society (WSIS) - Plan of Action <www.itu.int/wsis/docs/geneva/official/poa.html>

"Promote electronic publishing, differential pricing and open access *initiatives to make scientific information affordable and accessible in all countries on an equitable basis."[C7. E-science]*

b) UNESCO Universal Declaration on Cultural Diversity[1]

The United Nations-backed Universal Declaration on Cultural Diversity strongly promotes the concept of digitizing documentary heritage collections for initiating intercultural dialogue which is the best guarantee of peace (UNESCO, 2001). This Declaration confirms that member countries of the United Nations will take appropriate strategic steps to bring cross-cultural documentary heritage collections accessible to worldwide communities. An excerpt of relevant texts from both *Declaration* and Main Lines of an *Action Plan* is highlighted below:

"Towards access for all to cultural diversity While ensuring the free flow of ideas by word and image care should be exercised that all cultures can express themselves and make themselves known. Freedom of expression, media pluralism, multilingualism, equal access to art and to scientific and technological knowledge, including in digital form, and the possibility for all cultures to have access to the means of expression and dissemination are the guarantees of cultural diversity." [Article 6 of the Declaration]

"Encouraging "digital literacy" and ensuring greater mastery of the new information and communication technologies, which should be seen both as educational discipline and as pedagogical tools capable of enhancing the effectiveness of educational services. [Main Line # 9]

Promoting linguistic diversity in cyberspace and encouraging universal access through the global network to all information in the public domain. [Main Line# 10]

Countering the digital divide, in close cooperation in relevant United Nations system organizations, by fostering access by the developing countries to the new technologies, by helping them to master information technologies and by facilitating the digital dissemination of endogenous cultural products and access by those countries to the educational, cultural and scientific digital resources available worldwide." [Main Line # 11]

[1] UNESCO Universal Declaration on Cultural Diversity <http://unesdoc.unesco.org/images/0012/001271/127160m.pdf>

c) Open Access Statements and Declarations

In the wake of the open access movement, some policy frameworks have already been established by member communities to foster inclusive, plural and development-oriented knowledge societies. A number of open access declarations/ statements were made during the past decade, where the world's leading research institutions agreed on the open access mandates. An indicative list of major open access statements or declarations is given here:

- ARIIC Open Access Statement (Australian Research Information Infrastructure Committee) [www.caul.edu.au/scholcomm/OpenAccessARIICstatement.doc]
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities [http://oa.mpg.de/openaccess-berlin/berlindeclaration.html]
- Bethesda Statement on Open Access [www.earlham.edu/~peters/fos/bethesda.htm]
- Budapest Open Access Initiative Statement [www.soros.org/openaccess/]
- ERCIM Statement on Open Access (European Research Consortium for Informatics and Mathematics) [www.ercim.org/publication/Ercim_News/enw64/ercim-oa.html]
- IFLA Statement on Open Access to Scholarly Literature and Research Documentation [www.ifla.org/V/cdoc/open-access04.html]
- OECD Declaration on Access to Research Data from Public Funding (Organisation for Economic Co-operation and Development) [www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html]
- Washington DC Principles for Free Access to Science: A Statement from Not-for-Profit Publishers [www.dcprinciples.org/statement.htm]
- Wellcome Trust Position Statement in support of open and unrestricted access to published research [www.wellcome.ac.uk/doc_WTD002766.html]

d) Hyderabad Declaration, 2004

The ICT Ministers in the Asian Region met in Hyderabad in the *Asia IT Ministers' Second Summit* during 12-13 January 2004 and adopted a collective principle of undertaking definitive and firm steps to spread the benefits of ICT. In this

declaration, digitization of culture is one of the bold provisions to spread the benefits of ICT to the common citizens. An excerpt of relevant texts from the *Declaration* is highlighted below:

> ***Digitization of Culture***[1]
>
> *"This Summit recognizes the role of culture – that is, the entire spectrum of its heritage, values, practices, ways of life, knowledge systems, languages and artistic expression transmitted through generations - in promoting self-respect in communities and among nations. We realize the great significance of Asia as a centre of cultural diversity. Taking into consideration the fact that traditional expressions of culture are getting obscured, diluted and transformed, this Summit proposes to develop an Asia Digital Culture Community for developing regional cooperation, coordinating international collaboration and catalyzing knowledge preservation in member countries. The preservation of our collective cultures will be enhanced through such activities as:*
>
> - *Raising awareness among international communities and groups about the value of* digital documentation *of their culture.*
> - *Adopting an agreed code of ethics for collection, digital documentation and publication of cultural resources.*
> - *Developing local language technologies and on-line transmission capabilities.*
> - *Fostering training in conservation,* digitization *and dissemination.*
> - *Enabling IPR development for* free access and dissemination *of cultural resources."*

1.7.1.2 National Policy Instruments

a) NKC recommendations on Libraries and Open Educational Resources

The National Knowledge Commission of India (NKC), constituted on 13th June 2005, is a high-level advisory body to the Prime Minister of India, with a mandate to guide policy and direct reforms. NKC's overarching aim is to transform India into a vibrant knowledge-based society. The NKC's *Working Group on Libraries*

[1] Hyderabad Declaration - 2004. <http://asiaitsummit.nic.in/finalplan.htm>

and *Working Group on Open Access and Open Educational Resources* have strongly recommended digitization of documentary heritage collections, open access to public-funded research literature and supported establishment of open courseware repositories for countrywide dissemination of quality courseware to many cross sections of people. If implemented, these recommendations will have far-reaching implications in the knowledge creation and dissemination cycle. The documentary heritage collections, scholarly literature and lifelong learning materials digitized or produced by state-sponsored institutions would then be made accessible through open access channels such as national and institutional repositories. This way the NKC's recommendation on peer-reviewed research papers resulting from public-funded research would be validated by subject experts when making these resources available through open access channels. NKC also recommended creation of national knowledge portals for basic needs/ key sectors such as water, energy, environment, education, food, health, agriculture, employment and citizen rights. Already national portals on water (India Water Portal, www.indiawaterportal.org), energy (India Energy Portal, www.indiaenergyportal.org), environment (India Environment Portal www.indiaenvironmentportal.org.in) and education (Sakshat, www.sakshat.ac.in) have been launched that provide open access to information, knowledge and learning resources on the relevant areas. An excerpt of relevant recommendations and open access statements of NKC (2006, 2007) is highlighted below:

> ***Digitization and Open Access***[1]
>
> *"To enable equitable and universal access to knowledge resources, it is important to create more digital resources which can be shared. The concept of an 'information commons' i.e. 'resources shared by a community of producers and consumers in an open access environment' needs to be promoted. New resources should be openly accessible and historical documents, too, should be digitized and made available.*

[1] *Report of the Working Group on Libraries, National Knowledge Commission http://knowledgecommission.gov.in/downloads/documents/wg_lib.pdf*

- *This Working Group strongly recommends that peer-reviewed published research papers resulting from publicly funded research in India must be made available through* open access *channels, subject to copyright regulations. The group also recommends use of open standards and free and open source software.*
- *All pre-independence periodicals and newspapers in all Indian languages and in English must be digitized for access and preservation.*

To help preserve digital resources, optimize their use and avoid duplication of effort:

- *State-level archives for preservation of digitized materials must be set up.*
- *Every state should establish a registry and archives of knowledge-based digital resources, and make it accessible."*

Encourage Open Access[1]

"Open access material stimulates research and helps students, teachers and researchers across the world. Therefore at a policy level, all research articles published by Indian authors receiving substantial government or public funding must be made available under Open Access and should be archived in the standard OA format at least on his/her website. As a next step, a national academic OA portal should be developed. The government should allocate resources to increase the current digitization *efforts of books and periodicals which are outside copyright protection. Separate funding should be allocated to develop a new high quality OCR software package so that new and old fonts in many different Indian languages can be converted into ISCI/ASCI code and OA portals and servers could be upgraded regularly. Appropriate financial resources should be earmarked for these endeavours. This will also facilitate machine translation of these valuable resources."*

Summary of NKC Recommendations to Prime Minister of India[2]

On ***Libraries (2006)***

- *Set up a National Commission on Libraries*
- *Prepare a national census of all libraries*

[1] *Report of the Working Group on Open Access and Open Educational Resources, NKC http://knowledgecommission.gov.in/downloads/documents/wg_open_course.pdf*
[2] *Report to the Nation 2007*. New Delhi: National Knowledge Commission.

- *Revamp library information science (LIS) education, training and research facilities*
- *Re-assess staffing of libraries*
- *Set up a central library fund*
- *Modernize library management, encourage greater community participation in library management*
- *Promote information and communication technology (ICT) applications in all libraries*
- *Facilitate donation and maintenance of private collections*
- *Encourage public private partnerships in LIS development*

*On **Open Educational Resources (2007)***

- *Support the production of quality content by a select set of Indian institutions*
- *Leverage global open educational resources*
- *Encourage open access*
- *Develop network-enabled delivery infrastructure*
- *Create a faculty and institutional development programme*

*On **Other Thrust Areas – having relevance on Open Access***

- *Protection of traditional knowledge (TK) through Traditional Knowledge Digital Library (TKDL) and promoting incentives for wealth creation from TK [Intellectual Property Rights, 2007]*
- *Set up a National Education Foundation to develop web-based common open resources [Open and Distance Education, 2007]*

b) UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005[1]

In 2005, the University Grants Commission of India (UGC) drafted a national policy framework entitled "UGC (Submission of Metadata and Full-text of

[1] UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005 <http://www.ugc.ac.in/new_initiatives/etd_hb.pdf>

Doctoral Theses in Electronic Format) Regulations, 2005". This Regulation proposed two sets of planned actions such as:

- Creation of Indian National Theses Database (INTED)
- Submission of PhD Theses in Electronic Form

This set of regulations is still under consideration. They propose to formulate a roadmap to achieve widest dissemination of results of doctoral research conducted in Indian universities and bibliographic control of theses and dissertations of research degrees. So far, a few UGC-supported universities have established open access repositories for scholarly literature produced in the respective universities. Initially the universities were reluctant to change their status quo, as the contents of scholarly literature including the PhD theses will be a matter of critical analysis by the national and international peers, if the universities establish open access repositories. Now, universities will be motivated to produce qualitative doctoral theses and will maintain certain international standards. The UGC also supports development of infrastructure in Indian universities through various planned schemes. The proposed national education grid will also enrich modern ICT infrastructure in Indian universities. The Indian universities then will have necessary infrastructure to host a number of Web-based information services. Hosting an ETD repository and providing online interface to INTED will not be a problem in most of the universities. Indian universities should now take a proactive role in the implementation of these regulations at the earliest, ensure qualitative research and make the results of doctoral research widely available. The UGC should also clear all bottlenecks to implement INTED and decentralized ETD repositories across the country.

1.7.2 Digitization Guidelines and Best Practices

The international communities have developed several digitization guidelines and best practices to guide their member institutions on the effective ways to implement digitization projects. Some guidelines deal with technical specifications and other operational issues in great details, whereas some other

guidelines help the project managers in handling other aspects of the digitization projects such as human resources, finance, legal matters including intellectual property rights, advocacy, and how to involve stakeholders.

1.7.2.1 International Guidelines

Several guidelines are in existence at the national and international levels. Some international guidelines are followed in Indian digitization projects, whereas some national guidelines have adopted international guidelines with some degree of localization. Popular international guidelines are described in this Section.

a) *Guidelines for Digitization Projects for Collections and Holdings in the Public Domain, particularly those held by Libraries and Archives*

This document is jointly prepared and published by UNESCO, IFLA (International Federation of Library Associations and Institutions) and ICA (International Council on Archives) in March 2002. These Guidelines are the result of a joint venture of a group of experts on behalf of IFLA and ICA, who had been invited to draft these for UNESCO (2002). This document contains eight comprehensive chapters such as:

- Selection of documents for digitization
- Technical requirements & implementation
- Legal aspects
- Budgeting
- Human resource planning
- Development & maintenance of Web interfaces
- Preservation of digital content
- Project management.

b) *The Guide for Electronic Theses and Dissertations*

This guidebook is the result of a joint venture of a group of experts on behalf of UNESCO and NDLTD (Networked Digital Library of Theses and Dissertations). This document is prepared for ETD stakeholders such as students, researchers research supervisors, policymakers, ETD practitioners and NDLTD members (UENSCO, 2002). This guidebook provides a roadmap for implementing ETD at

institutional and national scale. Targeting different stakeholders, this guide contains several comprehensive chapters such as:

- Universities
- Students
- Technical issues
- Training the trainers
- The Future

c) *SPARC Institutional Repository Checklist & Resource Guide*

This is a comprehensive guidebook prepared by Raym Crow (2002) for the member institutions of the Scholarly Publishing & Academic Resources Coalition (SPARC). This guidebook helps the worldwide communities including Indian institutions in implementation of institutional repositories as well as open access subject repositories for wider dissemination of scholarly literature. The chapters in this document include:

- Securing Administration Support
- Securing Faculty Participation
- Librarians: Benefits and Challenges
- Repository Management and Policy Issues
- Technical and System Issues

1.7.2.2 National Guidelines

National level projects in India also prepared and published some guidelines based on international guidelines and best practices to achieve uniformity in digitization work carried out simultaneously at different digitization centres. Some important guidelines emanated from Indian digitization projects are:

a) Million Book Universal Library Project: Manual for Metadata Capture, Digitization, and OCR

This Manual was prepared by Gabrielle V. Michalek (2003) for the Digital Library of India and Universal Digital Library projects. This manual is used across scanning centres and regional mega scanning centres of DLI project in India.

This Manual covers operational aspects of mass digitization work for rare books and printed documents. This document contains five chapters, namely:

- Data Production
- Getting MARC Records from OCLC
- Creating Metadata Using Dublin Core
- Minolta PS 7000 – Quickscan Software Instructions
- ABBYY FineReader 6.0 Instructions

b) Guidelines for Digitization of Manuscripts

This document was prepared by the digitization team of National Mission for Manuscripts (2004). This manual is used by manuscript digitization centres of the Mission across India. This is a very comprehensive manual covering topics namely:

- Criteria for Selection
- Technical Requirements and Implementation
- Output Specification
- Metadata Creation
- Collection Management
- Image Quality Check List

c) UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005

This document was prepared in 2005 for national level consultation of stakeholders in Indian higher education system. This document provides a detailed roadmap for achieving ETD repositories at the national and university levels. This document is divided into two parts. The first part provides background information on Indian and international scenarios. In this part, some specific standards for theses data are also elaborated. The second part deals with implementation process. An outline of this manual is given below:

- Part I: Background Information, Current Scenario, Major Issues & Data Standards
    - Present Scenario in India

        - International Scenario
        - Standards for Data Format for Theses
    - Part II: Implementation Process
        - Creation of Metadata for Current Theses
        - Submission of Ph.D. Theses in Electronic Format
        - Setting-up Institutional Repositories for ETD
        - Steps for Setting-up Institutional Repositories
        - Technical Requirements

1.7.3 Literature Describing National Digitization Initiatives

Indian digital library and digitization initiatives are widely described and evaluated by the national and international researchers in their papers. Most of the projects described in the literature are successfully implemented national level initiatives (Vikas, 2005). Open access initiatives are also highlighted in some of the papers (Fernandez, 2006; Ghosh & Das, 2007; Rajasekhar, 2003; Suber & Arunachalam, 2005; Ahmed & Rather, 2007), describing different kinds of projects such as institutional repositories, subject repositories, open educational resources, open access journals, metadata harvesting services and library consortiums. Some of the literature communicated in journals and conferences can be classified as country report holistically describing country efforts in the areas of digital library development and open access movement. The several directories also provide comprehensive listing of open access projects implemented in India adhering to international standards, more particularly Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH). Examples of such directories are Directory of Open Access Journals (DOAJ), Directory of Open Access Repositories (Open DOAR), Registry of Open Access Repositories (ROAR), and Cross Archive Search Services for Indian Repositories (CASSIR). These directories have become effective information sources for several papers (Fernandez, 2006; Ghosh & Das, 2007; Mittal & Mahesh, 2008) describing country efforts presented in conferences and published in Indian journals.

Indian government agencies supporting digitization and open access projects across the country also provide comprehensive listings of completed and ongoing projects supported by them. For example, National Digital Libraries Cell of the Department of Information Technology in MCIT publishes a list of projects funded by MCIT and their current status including achievements[1]. Among the listed projects funded by MCIT, most prominent ones are Digital Library of India (DLI) and its associated projects, cultural heritage digital library projects at IGNCA. In this thesis both the initiatives has been dealt in great details.

Another agency supporting digitization and open access projects is the Technology Information Facilitation Programme (TIFP) under Department of Scientific & Industrial Research (DSIR), Government of India. This Programme is the successor of the National Information System for Science and Technology (NISSAT) programme of DSIR. TIFP publishes a list of digital information management projects funded by DSIR and their current status including achievements[2]. DSIR is also country focal point for the Asia Pacific Information Network (APIN), promoted by UNESCO's Information for All Programme (IFAP). Tuning with APIN's long-term strategies, TIFP's major thrust areas, emphasizing on content creation, digitization, digital library and open access development in the country, include:

- Indian digital library of theses and R&D publications;
- Electronic publishing of selected Indian S&T materials;
- Open archive initiatives – a Web alternative to scholarly communications;
- Documentation of traditional knowledge and folk wisdom;
- Promotion of information access and sharing;
- Promotion of content development;
- National websites/ servers.

[1] National Digital Libraries Cell (2008). List of Projects and Current Status including Achievements <www.mit.gov.in/default.aspx?id=325>

[2] Technology Information Facilitation Programme (2008). List of Projects Supported. <www.dsir.gov.in/tpdup/tifp/tifplop.htm>

Some of the successful digitization projects funded by DSIR and analyzed in this thesis are:

- Development of OAI-Based Institutional Research Repository Services in India, undertaken by National Centre for Science Information (NCSI), Indian Institute of Science.
- Vidyanidhi Digital Library, undertaken by University of Mysore.

1.7.3.1 International Conferences held in India dealing with Digital Libraries

Conference papers are major source of information on digital library initiatives in India. Many international conferences, several national conferences as well as national seminars held in India have produced a huge amount of literature greatly dealing with theoretical aspects of digitization, digital library development and open access archives. Only a limited portion of the conference literature deals with research problems as well as actually implemented cases. Few qualitative research papers presented in these conferences later got published in national and international peer-reviewed journals, whereas some other conference papers were archived in subject repositories such as Librarian's Digital Library and E-LIS. A list of international conferences on digital libraries held in India is given here:

- The Fourth International Conference of Asian Digital Libraries (ICADL 2001) was held in Bangalore during 10-12 December 2001. This conference was organized by University of Mysore and Indian Institute of Information Technology (IIIT) Bangalore. The theme of the Conference was Digital Libraries: Dynamic Landscape for Knowledge Creation, Dissemination and Management. ICADL is a series of international conferences with an Asian focus, first one held in Hong Kong in 1998. ICADL 2001 was the first major international conference held in India on digital libraries. In this conference, several Indian professionals presented papers mostly in the areas of theoretical understanding on the subject (Urs, Rajashekar & Raghavan, 2001).

- The First International Conference of Digital Libraries (ICDL 2004) was held in New Delhi during 24-27 February 2004. This conference was organized by TERI (The Energy and Resources Institute), New Delhi. The theme of the Conference was knowledge creation, preservation, access and management. This was the one of the major international conferences held in India on digital libraries. In this conference, over one hundred Indian professionals presented papers mostly in the areas of theoretical understanding on the subject. In some papers institutional digital library and digitization initiatives are also highlighted. The Conference proposed a National Mission on Digital Libraries in coalition with the other Missions initiated by the Department of Culture (Kar, 2004).
- The Second International Conference of Digital Libraries (ICDL 2006) was held in New Delhi during 5-8 December 2006. This conference was organized by TERI, New Delhi. The theme of the Conference was information management for global access. In this conference, over one hundred Indian professionals shared their practical experiences in the implementation of institutional digitization and institutional repository projects. The Conference proposed the enactment of a Digital Library Act to facilitate digitization and content creation for universal access to information (Kar, 2006).
- The First IEEE International Conference on Digital Information Management (ICDIM 2006), organized by the Digital Information Research Foundation Chennai and Chirst College Bangalore, was held in Bangalore during 6-8 December 2006. One of the themes of the Conference was digital libraries. In this conference, a number of professionals presented cases on Indian digital libraries and digital archives.
- The International Conference on Information Management in a Knowledge Society (ICIM 2005), organized by Indian Association of Special Libraries and Information Centres (IASLIC) to commemorate its golden jubilee celebrations, was held in Mumbai during 21-25 February 2005. One of the

themes of the Conference was digital libraries and digital archives. In this conference, a number of Indian professionals presented institutional cases on digital libraries and open access archives (Chandra, 2006).

- The International Conference on Semantic Web and Digital Libraries (ICSD 2007), organized by Documentation Research and Training Centre (DRTC), Indian Statistical Institute, was held in Bangalore during 21-23 February 2007. This Conference had two broad themes, viz., semantic Web and digital libraries and the convergence of both the technologies. In this conference, a number of Indian professionals presented theoretical frameworks on semantic Web concept and its integration with digital libraries and open access archives. Case studies of a few semantic Web applications in Indian open access and digital library initiatives were also presented in this Conference.

- The International Conference on Building Knowledge Repositories: Cross-Sectoral Collaborations, organized by National Institute of Fashion Technology (NIFT), was held in New Delhi during 7-9 February 2008. In this conference, a number of Indian professionals presented institutional cases on knowledge repositories and open access archives as well as shared their experiences on the cross-sectoral partnership in knowledge dissemination. The Conference recommended *(i) facilitating and enhancing cross-sectoral collaborations among institutions for developing digital repositories through a nodal agency; (ii) establishing a national depository for digital objects being produced by scholarly societies and institutions; and (iii) promoting open access standards for convergence and collaboration through research and training.*

1.7.3.2 Baseline Surveys and Country Reports on Digitization and Open Access Initiatives in India

A number of Indian researchers have published papers on state of the affairs of digitization and digital library initiatives in India. Some of the authors

contextualize the situations with their organizations they are working in their papers.

First beads of papers on the digitization issues were written by veteran library professionals citing their concerns on the implementation processes of this new domain. Dasgupta (2001, 2004) in her papers raises some concerns on the sustainability issue. To her viewpoint, sustainability is always in doubt as most of the digitization projects are started by getting a one time grant without a follow up plan. She points out that sustainability is a broad term which refers to many factors starting from technological issues of preserving digital data to social and economic questions for long-term accessibility of information to the people. For tackling sustainability factors, she suggests a focused attention on the broad spectrum comprising:

a) Policy and strategies,
b) Technical issues,
c) Economic aspects,
d) Social and educational issues,
e) Political and administrative will,
f) Capacity building,
g) Cooperative ventures, and
h) Changing role of librarians and information professionals.

Sood and Chandrasekharan (2004) published a baseline study of digital libraries in India, indicating different initiatives at the national and institutional level. In this paper technical and content creation issues are greatly dealt with. Kalra (2001), Singh (2002), Bhattacharya (2004), Gulati (2004), Jeevan (2004), Jain and Babbar (2006), Singh (2006), Mittal and Mahesh (2008) besides many others published baseline surveys in different years contextualizing technological development at that time. However, Das et al. (2007) analyzed search and retrieval features of eight Indian digital libraries available in public domain, namely, Digital Library of India, Kalasampada: Digital Library-Resource for Indian Cultural Heritage, Cultural Heritage Digital Library in Hindi, Traditional Knowledge Digital Library, Archives of Indian Labour: Integrated Labour History

Research Programme, Vidyanidhi Digital Library, Muktabodha Digital Library and Archiving Project, and e-Journals@INSA. Mittal and Mahesh (2008) also published literature review pertaining to digital library development in India, in addition to descriptive accounts of digital library initiatives in the country.

On the other hand, open access to information and knowledge is related concept to digitization although situated in different ideal plane. While digitization of library materials aims at long-term preservation of rare documents that are on the verge of extinction, open access to knowledge aims at providing access to current literature mostly available as born digital objects.

Mittal and Mahesh (2008), Ghosh and Das (2007), Ahmed and Rather (2007), Fernandez (2006), and Das et al. (2005) published baseline surveys on open access initiatives in India, indicating digital archiving software used for establishing open access archives. Ghosh and Das (2007) also analyzed reasons for India's leadership in open access movement and the present research situations in India in the changing global context. Fernandez (2006) received research grant from the Shastri Indo-Canadian Institute for the evaluative study of open access initiatives in India and he did some in-depth study on the subject emphasizing on existing policy frameworks versus policy requirements.

Arunachalam (2008, 2005) in his writings emphasizes on necessities of open access to scientific knowledge in building research capacity in developing countries including India. He also campaigns for the implementation of open access projects at the regional, national and institutional level. He publishes several articles on this topic in different journals and forums. Das (2008) publishes a book on open access to knowledge and information with a comprehensive directory of open access and digital library initiatives in South Asia. This book has comprehensive coverage on different kinds of initiatives such as digital library projects, open courseware, open access journals, metadata harvesting services, national level open access repositories and institutional repositories.

1.7.3.3 Institutional Case Studies on Digitization and Open Access Initiatives in India

In 2005, *Journal of Zhejiang University Science* published special proceedings issue of the first International Conference on Universal Digital Library (ICUDL 2005). In this issue several papers described research carried out in Digital Library of India project. Balakrishnan (2005) describes technological challenges and future research directions. Other papers describing research problems include: self healing information system for digital libraries (Prahllad and Black, 2005), text-to-speech (TTS) system for universal digital library (Ambati and Reddy, 2005), optical character recognition (OCR) system for Tamil text (Seethalakshmi et al., 2005), Om transliteration tool for Indian languages (Ganapathiraju et al., 2005), transliteration editors for Indian languages (Prahllad, 2005). *Vishwabharat@TDIL*, a journal of Technology Development for Indian Languages Programme also devoted a special issue on fonts technology and digital library of India. Coordinators of regional mega scanning centres in Digital Library of India project shared their project experience in this issue. Digital Library of India project is critically analyzed and described in Chapter two of this thesis.

Many other researchers have published institutional case studies describing their projects in different phases, from proposal stage to implementation phase. Arora (2003) provides a detailed account of South Asia's largest library consortia 'Indian National Digital Library in Engineering Science and Technology' (INDEST) from the initial proposal stage to implementation road maps. This paper also describes digital library and digitization activities carried out in Indian Institute of Technology Delhi, funded by different government agencies.

Mujoo-Munshi (2003) provides a detailed account of journals digitization project of Indian National Science Academy (INSA) titled e-Journals@INSA. This paper describes other digitization activities proposed and carried out in INSA, funded by different government agencies.

Indian Institute of Sciences (IISc), Bangalore is one of the pioneering institutions in India having successful open access initiatives. IISc open access initiatives

have become standard models for many institutional repositories in India and even aboard. Jayakanth (2008), Anuradha (2005), and Rajasekhar (2003) have published papers describing open access publishing model adopted in IISc for worldwide dissemination of research papers published by researchers and faculty members of IISc. They also described operational aspects of different open access initiatives in IISc such as EPrints@IISc, ETD@IISc and Pravabhi: IISc Publications database.

In two separate papers, Deb (2006) and Deb and Kar (2005) describe digitization activities and integration of digitized contents into existing hybrid digital library at The Energy and Resources Institute (TERI) in New Delhi.

In his paper, Sutradhar (2006) shares his experience on implementation of a intranet-base institutional repository in IIT Kharagpur using DSpace open source software.

In her paper, Patra (2006) shares her views on need for a digital library on ceramics. She also proposes a roadmap for implementing a digital library in ceramics in her institute.

National Library of India (NLI) is an institution of national importance under Ministry of Culture having rarest of collections of manuscripts and books. In 1999 NLI undertook a pilot project entitled "Down Memory Lane" to digitize its rare and brittle books. From February 1999 to June 2001, a total of 6601 books containing more than 2.5 million pages were scanned and archived in 548 CD-ROMs (in duplicate). This digitized collection can be consulted in NLI premises only. In a paper (National Library of India, 2003), NLI describes its digitization activities for manuscripts and rare books along with some technical details about the project.

There are many other institutional case studies mostly communicated in conferences and national journals (Doctor & Ramachandran, 2008; Bansode, 2008; Doctor, 2008; Doctor, 2007; etc.). In this section, only some significant cases are reviewed that have linkages with this research work.

1.8 Chapterization

This thesis addresses essentially two issues – the digitization of library materials available with memory institutions and digitization of scholarly materials produced by Indian researchers and making these scholarly materials available through open access channels. This section presents an overview as to how the various parts and sections of this research are related to the research problems and observations of the evidences leading to the recommendation of functional as well as business models for self-sustainability of the digitization projects. As indicated in Figure 1.1, each chapter, within the chapters 2 and 5, presents the evidence from perspectives of implemented projects. Chapter 2 presents critical appraisal of India's largest digitization initiative – Digital Library of India. Chapter 3 presents critical appraisal of digitization work pertaining to theses and dissertations – that makes the way to the creation of electronic theses and dissertations (ETD). In this chapter Indian initiatives such as Vidyanidhi Digital Library, ETD@IISc, OpenMed@NIC, and CSIR e-Thesis projects are described.

Chapter 4 presents critical appraisal of digitization work of documentary heritage collections. This chapter deals with digitization work for most venerable pieces of artifacts such as rare manuscripts, rare books, old photographs, photographic slides, audio tapes and microfilms. In this chapter, digitization initiatives of two prominent Indian institutions, namely, Indira Gandhi National Centre for the Arts (IGNCA) and National Mission for Manuscripts (NMM) have been evaluated.

Chapter 5 presents critical appraisal of digitization work of Indian periodicals, more specifically the projects carried out by the Indian National Science Academy (INSA), Indian Academy of Sciences (IAS), Indian Medlars Centre at the National Informatics Centre (medIND@NIC), Medknow Publications and Ramakrishna Math. Digitized journal collections make the way to the creation of open access journals in India. Thus, the genesis of open access journal development in India is also indicated in this chapter.

Chapter 6 concludes with the summary of findings and recommendation of futuristic functional models suitable for changing contexts in the country.

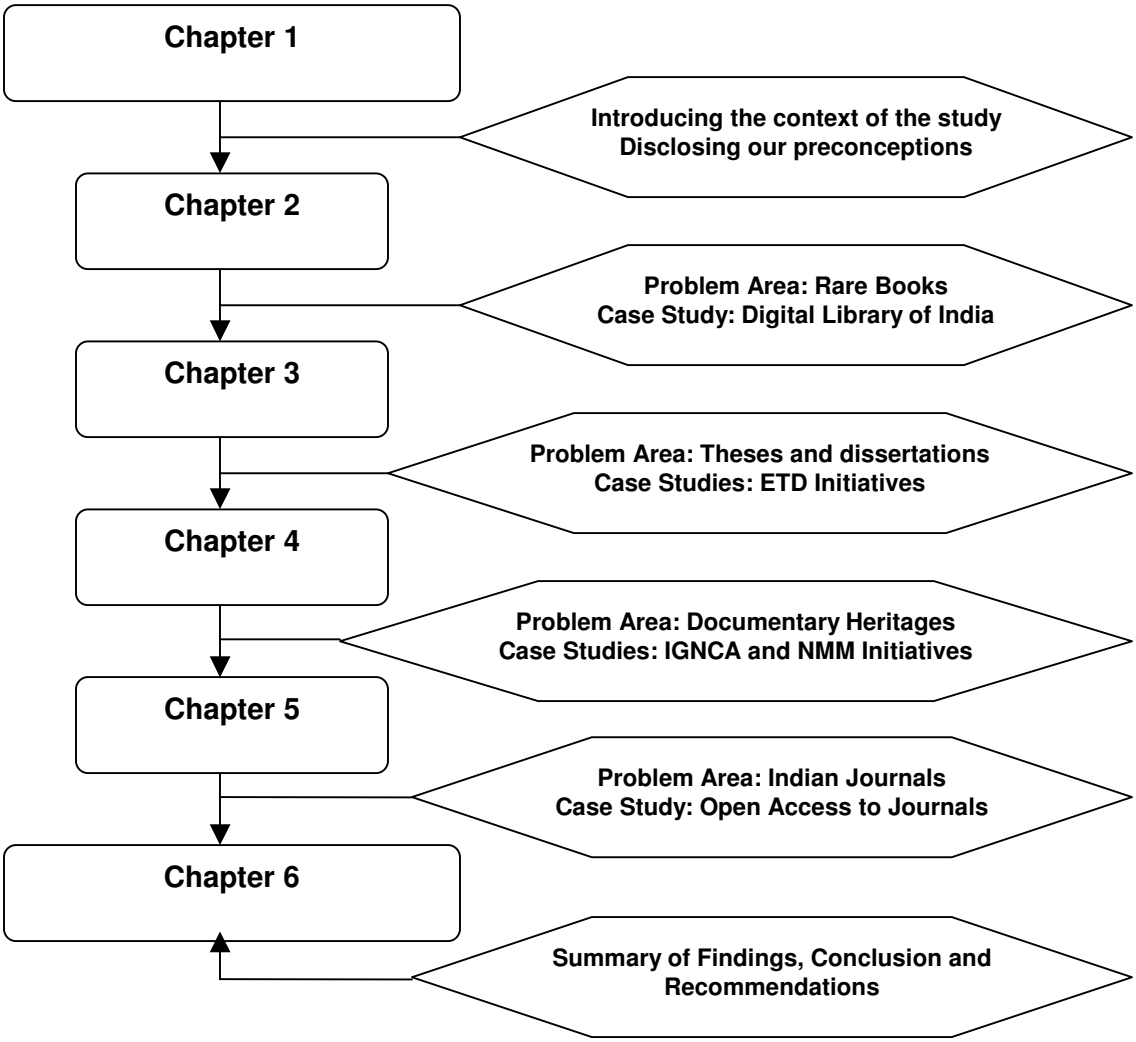


Fig. 1.1: Research Design

1.9 References

Ahmed, F. & Rather, R. (2007). Open access digital repositories: an Indian scenario. *KnowGenesis International Journal for Technical Communication*, 2(1), 16-20.

Ambati, V. & Reddy, R. (2005). Towards a self healing information system for digital libraries. *Journal of Zhejiang University Science*, 6A(11),1221-1228.

American Psychological Association (2001). *Publication manual of the American Psychological Association*, 5th edition. Washington, DC: American Psychological Association.

Anuradha, K.T. (2005). Design and development of institutional repositories: A case study. *The International Information & Library Review*, 37(3), 169-178

Arora, J. (2003). Indian National Digital Library in Engineering Science and Technology (INDEST): a proposal for strategic cooperation for consortia-based access to electronic resources. *The International Information & Library Review*, 35(1), 1-17.

Arunachalam, S. (2008). Open access to scientific knowledge. *DESIDOC Journal of Library and Information Technology*, 28(1), 7-14.

Balakrishnan, N. (2005). Universal Digital Library – future research directions. *Journal of Zhejiang University Science*, 6A(11),1204-1205.

Balakrishnan, N., Reddy, R., Ganapathiraju, M. & Ambati, V. (2006). Digital Library of India: a testbed for Indian language research. *TCDL Bulletin*, 3(1).

Bansode, S. (2008). Creation of digital library of manuscripts at Shivaji University, India. *Library Hi Tech News*, 25(1), 13-15.

Benner, M. (2003). The digital archive of the Swedish East India Company, 1731-1813: a joint project of a university library and a history department. *Online Information Review*, 27(5), 328-332.

Bhattacharya, P. (2004). Advances in digital library initiatives: a developing country perspective. *The International Information & Library Review*, 36(3), 165-175.

Chandra, H. (2006). Indian Association of Special Libraries and Information Centers golden jubilee all India conference report. *Library Hi Tech News,* 23(3), 3-4.

Crow, R. (2002). *SPARC institutional repository checklist & resource guide*. Washington, DC: The Scholarly Publishing & Academic Resources Coalition. Retrieved 15 July 2008 from www.arl.org/sparc/bm~doc/IR_Guide_&_Checklist_v1.pdf

Das, A.K. & Mukherjee Das, A. (2008). Role of lifelong learning in emerging knowledge economy in India. In: *Prof Alaka Buragohain Festschrift Volume*. Guwahai: Assam College Librarians' Association.

Das, A.K. (2008). *Open access to knowledge and information: scholarly literature and digital library initiatives – the South Asian scenario*. New Delhi: UNESCO.

Das, A.K., Dutta, C. & Sen, B.K. (2007). Information retrieval features in Indian digital libraries: a critical appraisal. *OCLC Systems & Services*, 23(1), 92-104.

Dasgupta, K. (2001). Asian content development: unity in diversity. *Library Review*, 50(7/8), 403-406.

Dasgupta, K. (2005). Digitization, sustainability and access in the Indian context. *Presented in the World Library and Information Congress: 71st IFLA General Conference and Council, "Libraries - A voyage of discovery"*, 14-18 August 2005, Oslo, Norway. Retrieved 15 July 2008 from www.ifla.org/IV/ifla71/papers/132e-Dasgupta.pdf

Deb, S. & Kar, D.C. (2005). Setting up an electronic library: the case of TERI. *The Electronic Library*, 23(2), 189-199.

Deb, S. (2006). TERI Integrated Digital Library Initiative. *The Electronic Library*, 24(3), 366-379.

Doctor, G. & Ramachandran, S. (2008). DSpace@IBSA: knowledge sharing in a management institute. *VINE*, 38(1), 42-52

Doctor, G. (2007). Knowledge sharing: developing the digital repository of SIPS. *VINE*, 37(1), 64-73.

Doctor, G. (2008). Capturing intellectual capital with an institutional repository at a business school in India. *Library Hi Tech*, 26(1), 110-125.

Fernandez, L. (2006). Open access initiatives in India - an evaluation. *Partnership: the Canadian Journal of Library and Information Practice and Research*, 1(1).

Ganapathiraju, M., Balakrishnan M., Balakrishnan N. & Reddy R. (2005). Om: One tool for many (Indian) languages. *Journal of Zhejiang University Science*, 6A(11),1348-1353.

Gaur, R.C. (2003). Rethinking the Indian digital divide: the present state of digitization in indian management libraries. *The International Information & Library Review*, 35(2-4), 189-203.

Ghosh, S.B. & Das, A.K. (2007). Open access and institutional repositories - a developing country perspective: a case study of India. *IFLA Journal*, 33(3), 229-250.

Gulati, A. (2004). Use of information and communication technology in libraries and information centres: an Indian scenario. *The Electronic Library*, 22(4), 335-350.

Jain, P.K. & Babbar, P. (2006). Digital libraries initiatives in India. *The International Information & Library Review*, 38(3), pp. 161-169

Jain, P.K. (2003). Indian Council of Social Science Research (ICSSR) maintained research institutes' libraries in India: towards digitization and networking. *The International Information & Library Review*, 35(2-4), 217-232.

Jayakanth, F. et al. (2008). ePrints@IISc: India's first and fastest growing institutional repository. *OCLC Systems & Services*, 24(1), 59-70.

Jeevan, V.K.J. (2004). Digital library development: identifying sources of content for developing countries with special reference to India. *The International Information & Library Review*, 36(3), 185-197.

Kalra, H.P.S. (2001). Efforts towards digitization of libraries in india: problems and prospects. *The International Information & Library Review*, 33(2-3), 197-204.

Kar, D.C. et al., eds. (2006). *Information management for global access: Proceedings of the 2nd International Conference of Digital Libraries* (ICDL 2006). New Delhi: The Energy and Resources Institute.

Kar, D.C. et al., eds. (2004). *Knowledge creation, preservation, access and management: Proceedings of the International Conference of Digital Libraries* (ICDL 2004). New Delhi: The Energy and Resources Institute.

Kotler, P. & Armstrong, G. (2003). *Principles of marketing*, 11th edition. New Delhi: Prentice Hall of India.

Krishnamurthy, M. (2005). Digital library of mathematics using DSpace: a practical experience. *SRELS Journal of Information Management*, 42(3), 245-256.

Krishnamurthy, M. (2005). Digital library services at the Indian Statistical Institute. *The Electronic Library*, 23(2), 200-203.

Michalek, G.V., ed. (2003). *Million Book Universal Library Project: manual for metadata capture, digitization, and OCR*. Carnegie: Carnegie Mellon University. Retrieved 15 May 2008 from www.library.cmu.edu/Libraries/MillionBookManual.pdf

Ministry of Communications and Information Technology (2004). *Digitization of culture - background paper for Asia IT Ministers' 2nd Summit*. Retrieved 15 July 2008 from http://asiaitsummit.nic.in/Digitisation.pdf

Ministry of Communications and Information Technology (2004). *Hyderabad Declaration 2004*. Retrieved 15 July 2008 from http://asiaitsummit.nic.in/finalplan.htm

Mittal, R. & Mahesh, G. (2008). Digital libraries and repositories in India: an evaluative study. *Program: electronic library and information systems*, 42(3), 286-302.

Moghaddam, G.G. & Talawar, V.G. (2008). The use of scholarly electronic journals at the Indian Institute of Science: a case study in India. *Interlending & Document Supply*, 36(1), 15-29.

Moxley, J.M., Masiello, D. & Fox, E.A. (2002). *The guide for electronic theses and dissertations*. Paris: UNESCO. Retrieved 15 July 2008 from http://etdguide.org/etdguide.pdf

Mujoo-Munshi, U. (2003). Building digital resources: creating facilities at INSA. *The International Information & Library Review*, 35(2-4), 281-309.

National Digital Libraries Cell (2008). *List of projects and current status including achievements*. Retrieved 15 July 2008 from www.mit.gov.in/default.aspx?id=325.

National Knowledge Commission (2006). *Report of the Working Group on Libraries*. New Delhi: National Knowledge Commission.

National Knowledge Commission (2006). *Report to the Nation 2006*. New Delhi: National Knowledge Commission.

National Knowledge Commission (2007). *Report to the Nation 2007*. New Delhi: National Knowledge Commission.

National Library of India (2003). Digitization of Manuscripts of the National Library of India. *CDNLAO Newsletter*, 47. Retrieved 15 July 2008 from www.ndl.go.jp/en/publication/cdnlao/047/473.html.

National Mission for Manuscripts (2004). *Guidelines for digitization of manuscripts*. Retrieved 15 July 2008 from http://namami.gov.in/pdfs/final%20digitization.pdf

Pandian, M.P., Jambhekar, A. & Karisiddappa, C.R. (2002). IIM digital library system: consortia-based approach. *The Electronic Library*, 20(3), 211-214.

Patra, C. (2006). Developing a digital library on ceramics. *The Electronic Library*, 24(4), 453-469.

Prahallad, K. & Black, A. (2005). A text to speech interface for Universal Digital Library. *Journal of Zhejiang University Science*, 6A(11), 1229-1234.

Prahallad, L., Prahallad, K. & Ganapathiraju, M. (2005). A simple approach for building transliteration editors for Indian languages. *Journal of Zhejiang University Science*, 6A(11),1354-1361.

Rajasekhar, T.B. (2003). Improving the Visibility of Indian Research: An Institutional Open Access Publishing Model. Presented in *Indo-US.Workshop on Open Digital Libraries and Interoperability 2003*. Retrieved 15 July 2008 from http://fox.cs.vt.edu/IndoUSdl/raja.pdf.

Reitz, J. M. (2008). *ODLIS - Online dictionary for library and information science*. Retrieved 15 July 2008 from http://lu.com/odlis/

Seethalakshmi, R. et al. (2005). Optical Character Recognition for printed Tamil text using Unicode. *Journal of Zhejiang University Science*, 6A(11), 1297-1305.

Shukla, V.N., Arora, K.K. & Gugnani, V. (2004). *Digital library: language centered research, test beds and applications.* Retrieved 15 May 2008 from www.cdacnoida.in/technicalpapers/PaperICDL1.pdf

Singh, J. (2002). From atoms to bits: consequences of the emerging digital divide in India. *The International Information & Library Review*, 34(2), 187-200.

Singh, S.P. (2008). Digitization of libraries: some Indian initiatives. Presented in *International Conference for Information and Knowledge Management 2008*, Kathmandu, Nepal. Retrieved 15 July 2008 from www.healthnet.org.np/ocs/index.php/icikm-2008/icikm/paper/view/15/12/.

Sood, A.D. & Chandrasekharan, U. (2004). *Digital libraries in India: a baseline study*. Bangalore: Educational Development Centre.

Suber, P. & Arunachalam, S. (2005). Open access to science in the developing world. In: *Open access: unrestricted access to publisher research*. Hyderabad: 93rd Indian Science Congress.

Sutradhar, B. (2006). Design and development of an institutional repository at the Indian Institute of Technology Kharagpur. *Program: electronic library and information systems*, 40(3), 244-255.

Swan, Alma (2008). Open access for Indian scholarship. *DESIDOC Journal of Library and Information Technology*, 28(1), 15-24.

Technology Information Facilitation Programme (2008). *List of projects supported.* Retrieved 15 July 2008 from www.dsir.gov.in/tpdup/tifp/tifplop.htm.

Tirumala Tirupati Devasthanams (2005). Tirumala Tirupati Devasthanams SV Digital Library. *Herald of Library Science*, 44(1/2), 77-81.

UNESCO (2001). *Universal Declaration on Cultural Diversity*. Retrieved 15 July 2008 from http://unesdoc.unesco.org/images/0012/001271/127160m.pdf

UNESCO (2002). *Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives*. Retrieved 15 July 2008 from www.ifla.org/VII/s19/pubs/digit-guide.pdf.

UNESCO (2002). *UNESCO guide to electronic theses and dissertations*. Retrieved 15 July 2008 from http://etdguide.org/etdguide.pdf

UNESCO (2003). *Digital libraries in education: analytical survey*. Retrieved 15 July 2008 from http://unesdoc.unesco.org/images/0013/001310/131025eb.pdf.

United Nations (2006). *WSIS outcome documents booklet*. Geneva: International Telecommunication Union.

Urs, S.R., Rajashekar, T.B. & Raghavan, K.S., eds. (2001). *Digital libraries: dynamic landscapes for knowledge creation, access and management* - Proceedings of the 4th International Conference of Asian Digital Libraries (ICADL 2001). Mysore: University of Mysore.

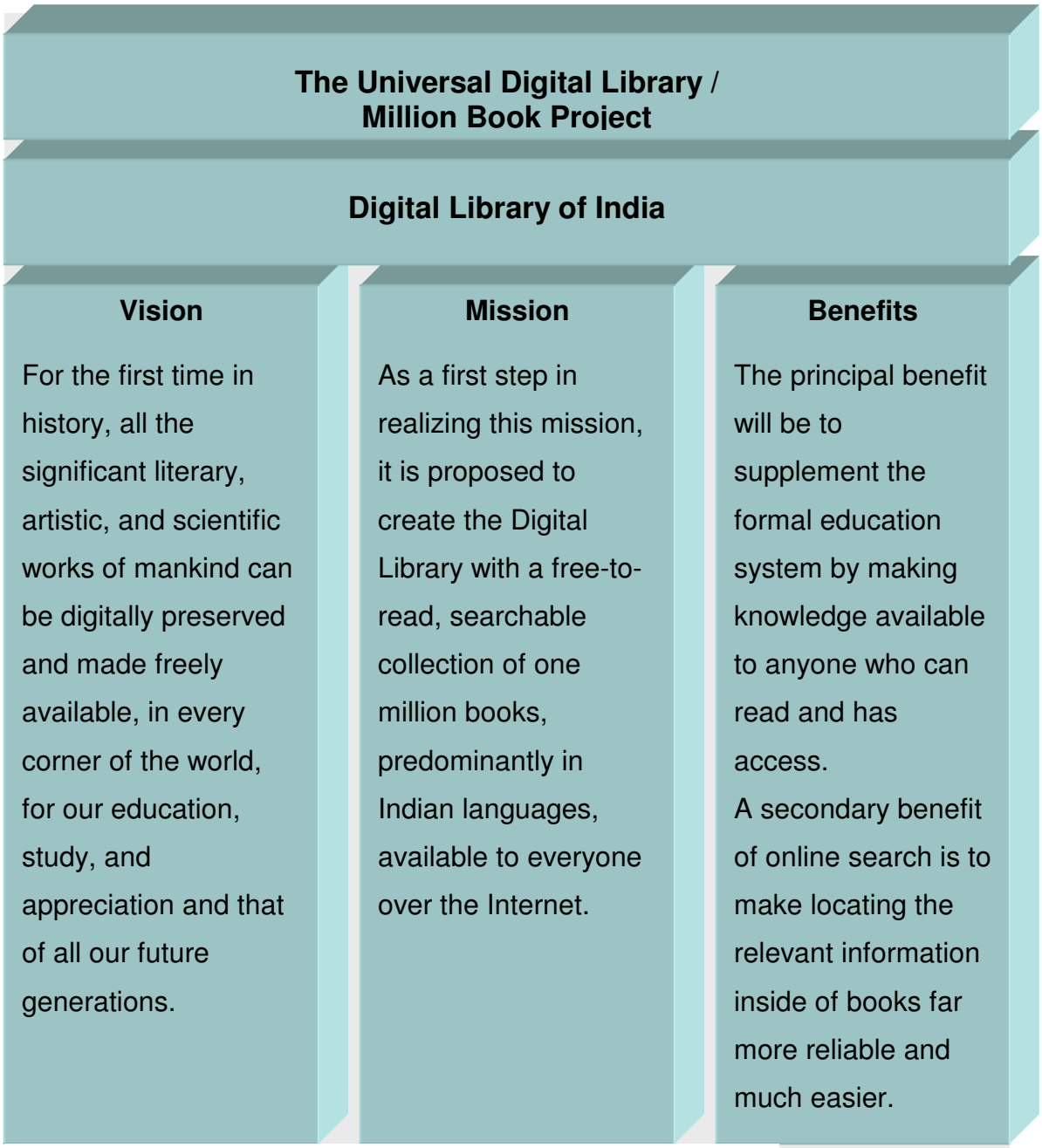
Vikas, Om, ed. (2005). Special issues on fonts technology and Digital Library of India. *Vishwabharat@TDIL*, 17, 59-106.

CHAPTER 2

India's Largest Digitization Initiative – Digital Library of India

The Digital Library of India (DLI) is the large scale digital library initiative in South Asia, spreading all over the country and establishing a network of four mega scanning centres and 21 scanning centres, which feed the digital contents into the digital library systems. Dr. A.P.J. Abdul Kalam, the former President of India, launched the portal of Digital Library of India on 8 September, 2003. Digital Library of India project is a part of the Universal Digital Library (UDL) or Million Book Project, coordinated by the Carnegie Mellon University in United States of America. Although the original UDL project has already been concluded and successfully implemented in the United States, the DLI project still ripples around with much acclamation. The former President of India, Dr. Abdul Kalam has shown keen personal interest for overall development and implementation of this project. In different forums, he promotes Digital Library of India for the benefits of younger generations and nation building. He also provided digital contents of his own writings for hosting in the DLI portals.

The Vision and Mission statements of DLI are similar to the same of parent Universal Digital Library project, except some local variations. Figure 2.1 depicts the Vision and Mission statements of Digital Library of India project. This is the simplified version of Vision and Mission statements as elaborated in the DLI portals. Figure 2.1 also indicates two major benefits of Digital Library of India project, i.e. (i) To supplement the formal education system by making knowledge available to anyone who can read and has access; and (ii) to make locating the relevant information inside of books far more reliable and much easier.



Source: http://www.new.dli.ernet.in/pturn/book_dli.html

Fig. 2.1: Vision and Mission Statements of Digital Library of India

2.1 Positioning Digital Library of India Initiative in Global Context

The Digital Library of India project is an essential part of Universal Digital Library project that aims at providing access to million plus books. Universal Digital Library is a global initiative of four nations, namely, the United States of America, India, China and Egypt, for disseminating centuries-old indigenous knowledge of mankind, spreading in the sub-contents of the countries.

Figure 2.2 illustrates the international dimension of UDL project, with an indication of availability of associated UDL portals that disseminate million plus digitized books in different languages. Figure 2.2 also specifies that Digital Library of India initiative is a subset of Universal Digital Library project. Other subsets of the UDL project are (i) China-US Million Book Digital Library Project, hosted by Zhejiang University, China (ii) Digital Assets Repository, hosted by Bibliotheca Alexandrina, Egypt and (iii) The Universal Library, hosted by Carnegie Mellon University, USA. Each subset of this global project is responsible for digitizing important literature produced in the respective country or sub-continent. A vast amount of oriental literature in different Indian languages produced in Indian sub-continent is thus digitized and hosted in DLI portals. National portals of DLI have country-specific identity, similar to other three member countries.

Figures 2.3 and 2.5 to 2.10 provide glimpse of the seven portals associated with the UDL project. Figure 2.3 provides a screen snapshot of UDL main portal. This Figure also indicates that this portal has three mirror sites in China North, China South and India, although China South was only functional during the study. UDL main portal provides quick and advanced search facilities with a single of metadata or combination of metadata. Searchable metadata available with this portal are: Title, Author, Subject, Language, Year and Country. The collections in this portal can be navigated through title of the books, author's last name, a range of years, subject and language. This portal is hosted by Carnegie Mellon University in USA. This portal is a conglomeration of collections of all affiliated UDL projects. Figure 2.4 indicates the contribution of different countries in the UDL portal in terms of available digitized books. This shows UDL portal has

received maximum number of digitized books from Chinese partner (75.48%), whereas Indian partner stands second (22.21%) and Egyptian partner third (2.31%). This pie chart also discloses that Chinese, Indian and Egyptian partners of UDL are solely responsible for content creation through large-scale digitization initiatives at the national level, whereas UDL partner in USA is responsible for strategic planning and hosting of digitized contents.

Carnegie Mellon University (CMU) hosts another portal of UDL project, which is nothing but almost mirror site of UDL international portal. Figure 2.5 provides a screen snapshot of the Universal Library portal, hosted by Carnegie Mellon University. On the homepage of this portal, Vision statement of UDL project is given along with the quick and advanced search interfaces.

Founded in 1897, Zhejiang University is one of China's oldest and most prestigious institutions of higher education. Zhejiang University hosts China-US Million Book Digital Library Project, which is the Chinese partner of UDL project. Figure 2.6 provides a screen snapshot of the portal for China-US Million Book Digital Library Project. On the homepage of the portal, quick search interface is given. From this search interface, a user can limit his search to documents pertaining to journals, dissertations, paintings, or videos, and English language documents.

Bibliotheca Alexandrina, the national library of Egypt, is a wonderful reincarnation of the famed ancient library of Alexandria. Bibliotheca Alexandrina is the Egyptian partner of UDL project and hosts Digital Assets Repository (DAR), another partner site of UDL. Homepage of DAR (Figure 2.7) provides quick search facility, quick links to eleven DAR categories and information on latest books.

In India, UDL project partners maintain three associated portals in the name of Digital Library of India sites. Figures 2.8 to 2.10 provide screenshots of DLI portals, which are hosted in three different locations having different sets of collections. Details of these three portals are elaborated in the subsequent sections of this chapter.

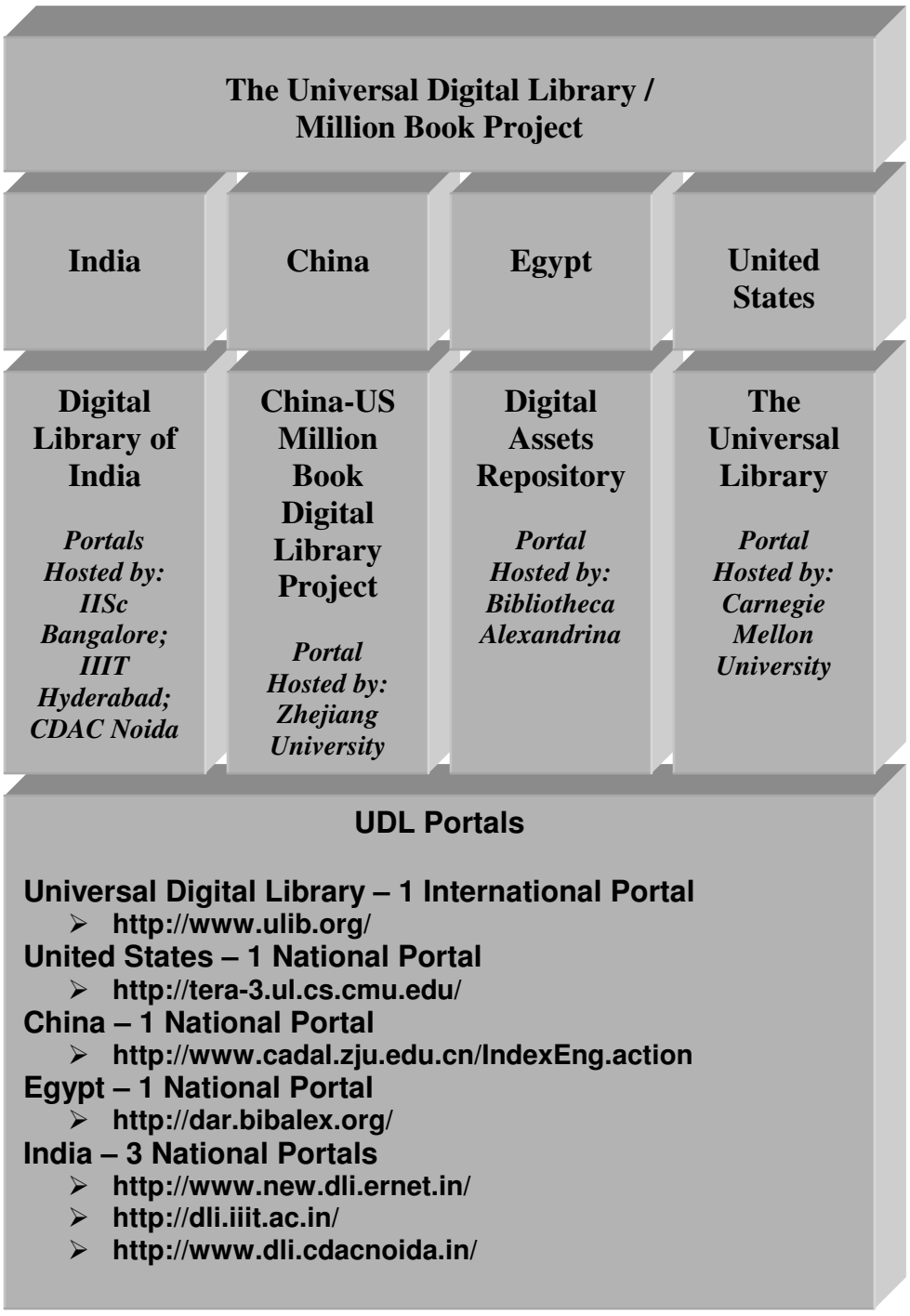
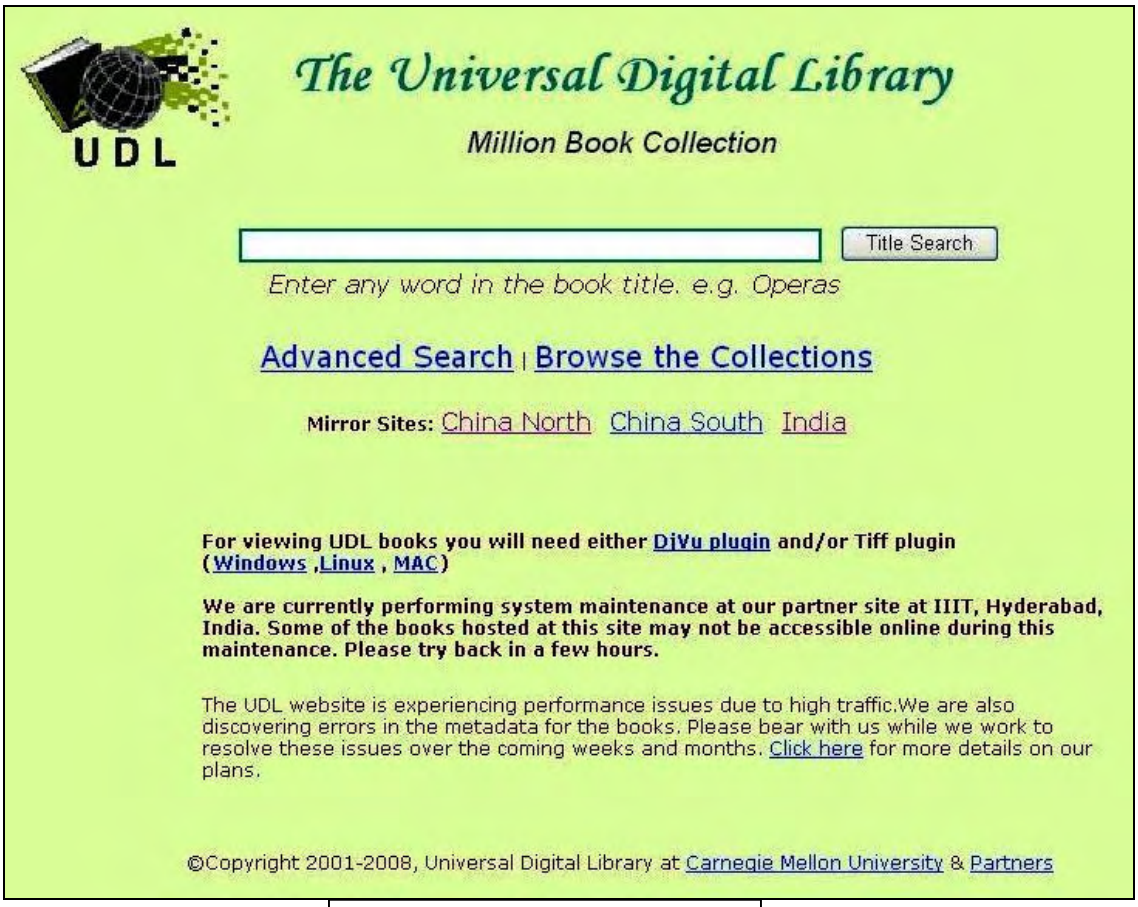


Fig. 2.2: Mapping of International Partners of UDL Project



http://www.ulib.org/

Fig. 2.3: Homepage of the Universal Digital Library Portal

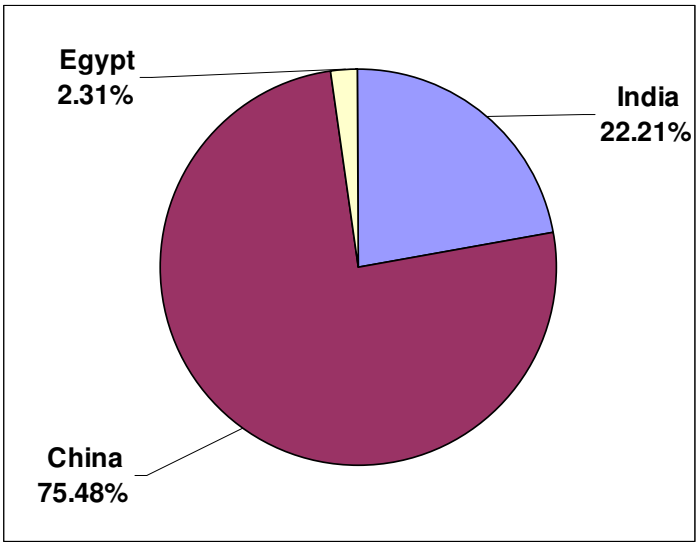


Fig. 2.4: Country-wise Contribution of Digitized Books in the UDL Portal

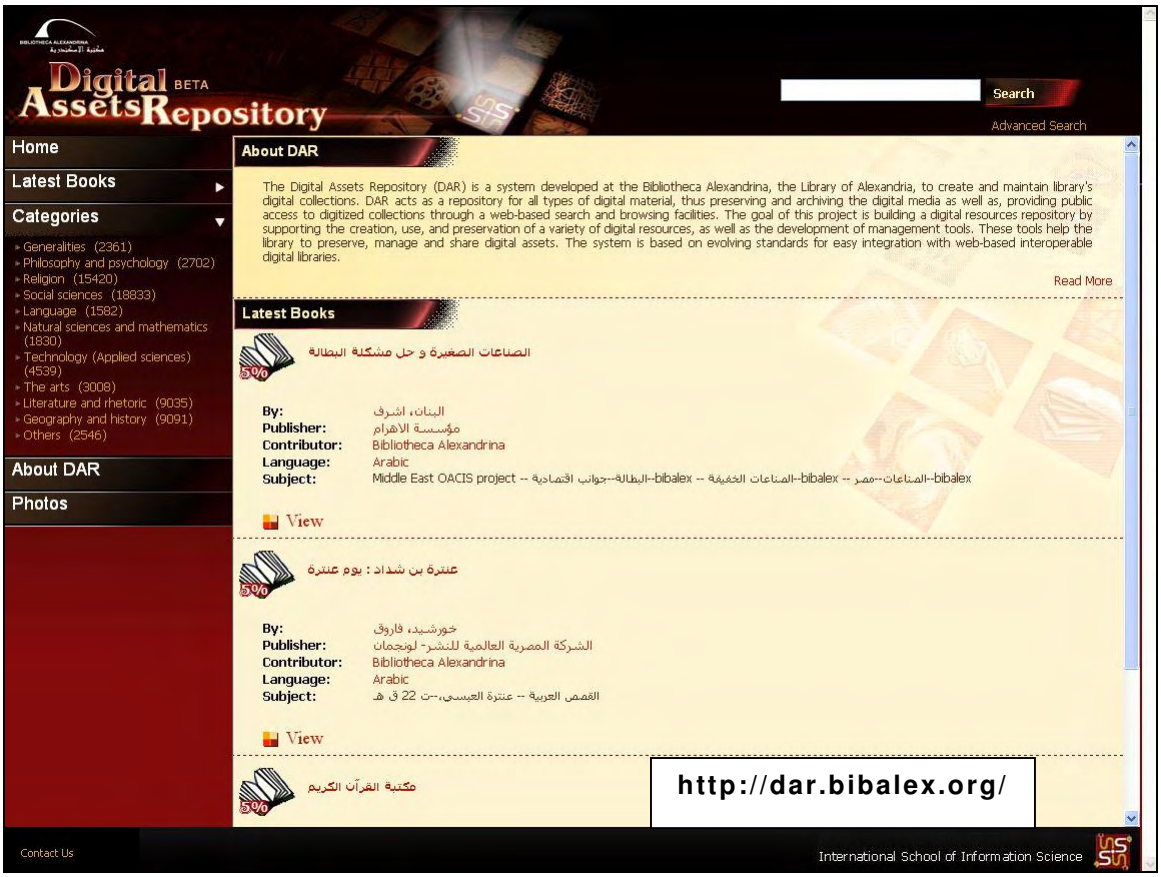
(As on 3 July 2008)

http://tera-3.ul.cs.cmu.edu/

Fig. 2.5: Homepage of the UDL Portal of CMU

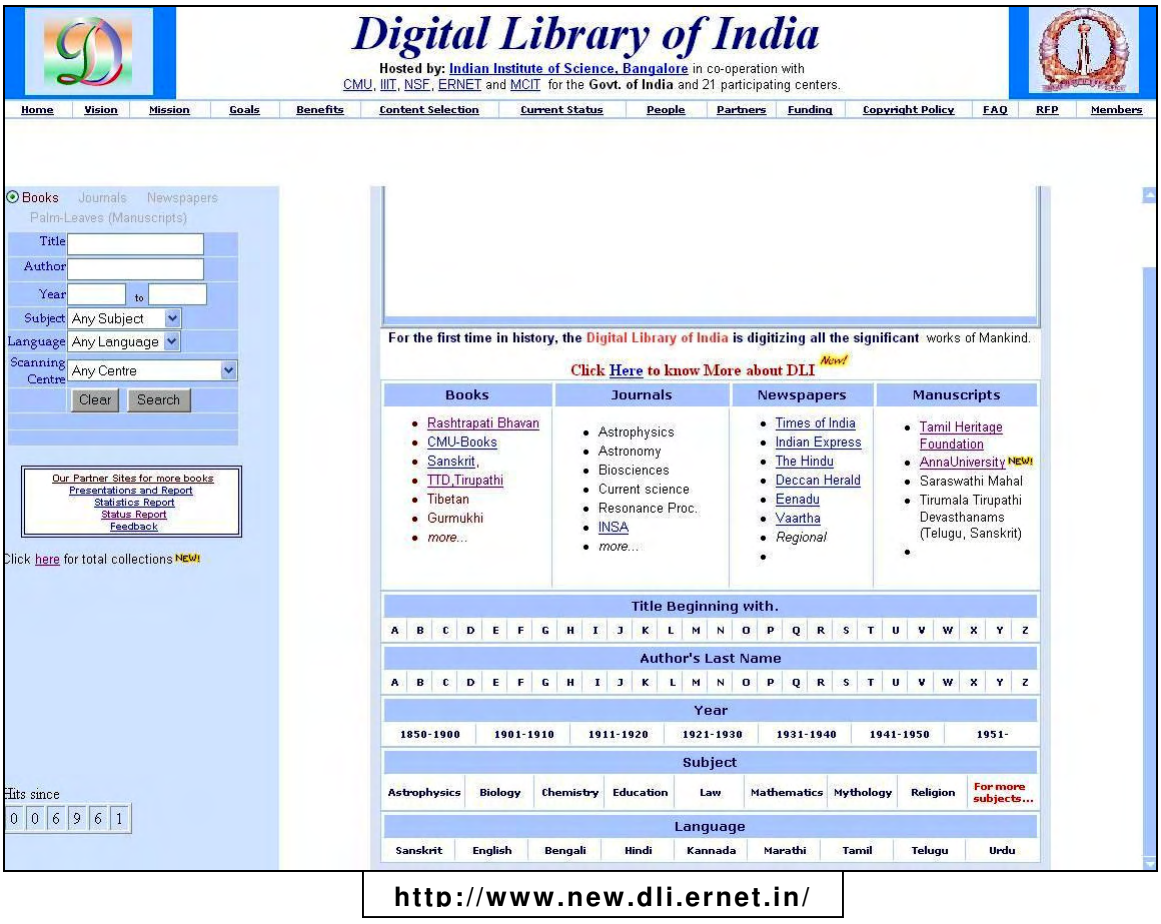
http://www.cadal.zju.edu.cn/IndexEng.action

Fig. 2.6: Homepage of the UDL Portal of China-US Project



http://dar.bibalex.org/

Fig. 2.7: Homepage of the UDL Portal of Egypt



http://www.new.dli.ernet.in/

Fig. 2.8: Homepage of the UDL Portal of IISc Bangalore

http://dli.iiit.ac.in/

Fig. 2.9: Homepage of the UDL Portal of IIIT Hyderabad

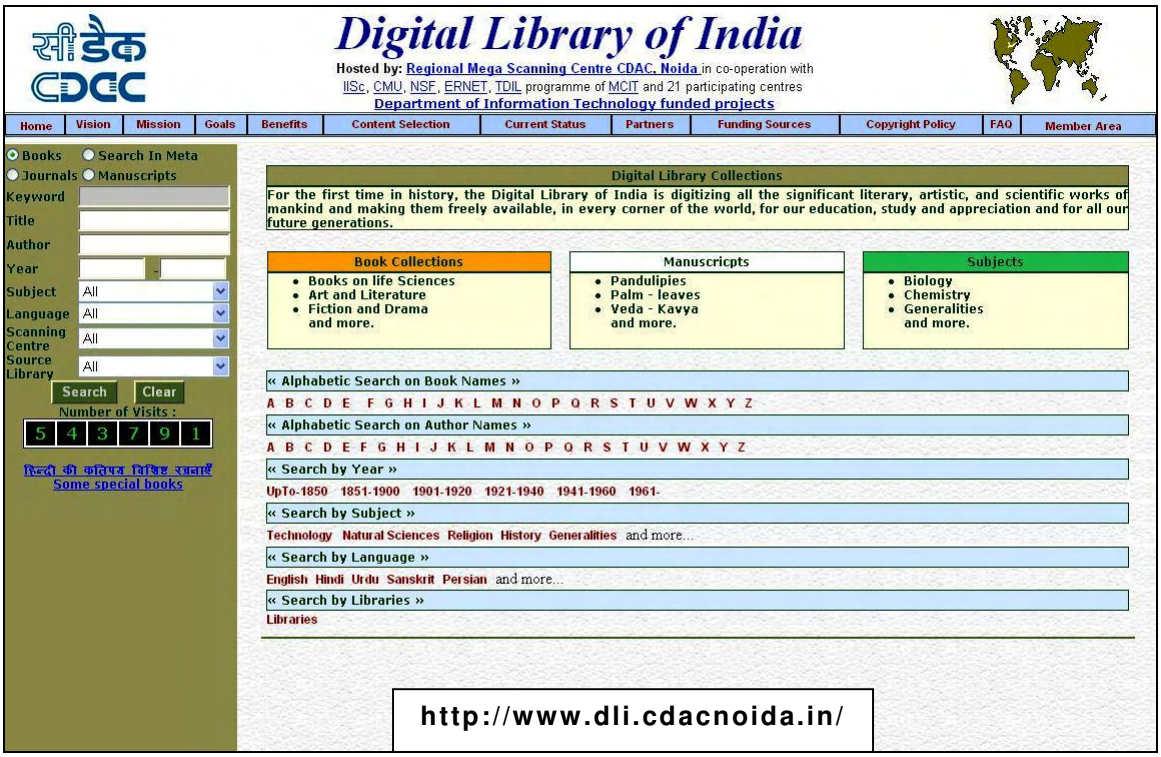
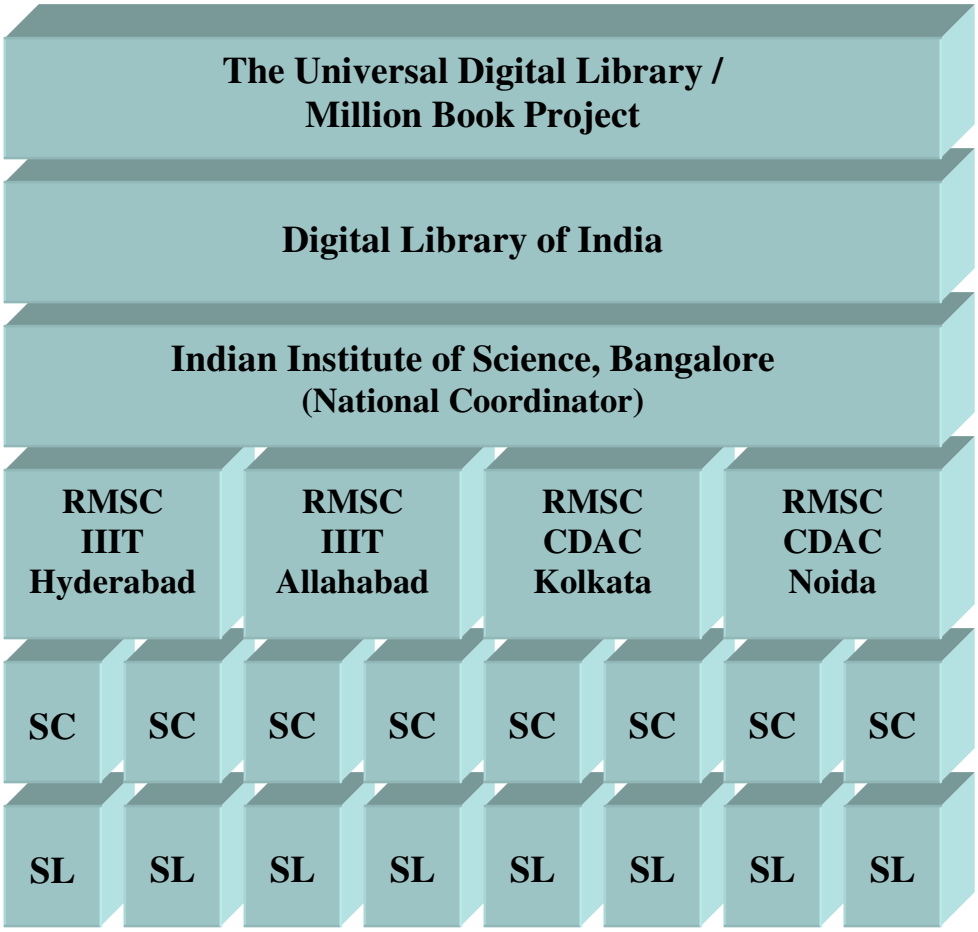


Fig. 2.10: Homepage of the UDL Portal of CDAC Noida

2.2 Structure of Digital Library of India Initiative

Digital Library of India was initiated at the national scale to digitize significant collections of old and rare documents available with important libraries and cultural institutions across India. DLI project maintains a decentralized structure, where Indian Institute of Science Bangalore (IISc) acts as National Coordinator of this project and as international collaborator with the UDL partners abroad. IISc is also a research partner of UDL project, responsible for development of applications software related to the project, in close collaboration with another research partner International Institute of Information Technology Hyderabad (IIIT Hyderabad). For maintaining whole country coverage, DLI project established four regional mega scanning centres across India, three in different metropolitan cities and one in a big city. Regional mega scanning centres (RMSCs) also maintain decentralized structure where local scanning centres carry out digitization activities at their end, in partnership with source libraries situated in

close proximity of the respective scanning centre. Figure 2.11 illustrates a schematic structure of Digital Library of India project. This picture also depicts functional hierarchy of each individual entity within the whole DLI partner institutions. Rolling of responsibilities can also be visualized in this Figure as National Coordinator IISc takes care of activities in the RMSCs. Similarly, each RMSC coordinates activities in local scanning centres, and each scanning centre secures continuous supply of physical documents from source libraries. On the other hand, captured digitized documents are being transferred from scanning centres to respective regional mega scanning centre in the specified formats. Then RMSC undertakes rigorous quality checking processes and uploads both metadata and digitized contents into the respective DLI portal. Regional Mega Scanning Centres in Kolkata and Allahabad again transfer captured digitized contents to DLI national office in Bangalore for uploading into DLI portal in IISc. The ERNET India is the backbone of DLI project, providing ICT infrastructural supports to the DLI partner institutions, including high storage Web servers and high speed broadband Internet connectivity. Figure 2.12 illustrates a layer of ICT connectivity provided by ERNET India that makes out a backbone to the DLI project. Having a dedicated broadband connectivity provided by ERNET India in different centres of DLI project, data flow from one level to next higher level becomes robust as shown in Figure 2.13. This Figure also depicts data transfer process within DLI partner institutions.



RMSC = Regional Mega Scanning Centre
SC = Scanning Centre
SL = Source Library

Fig. 2.11: Schematic Structure of Digital Library of India Project

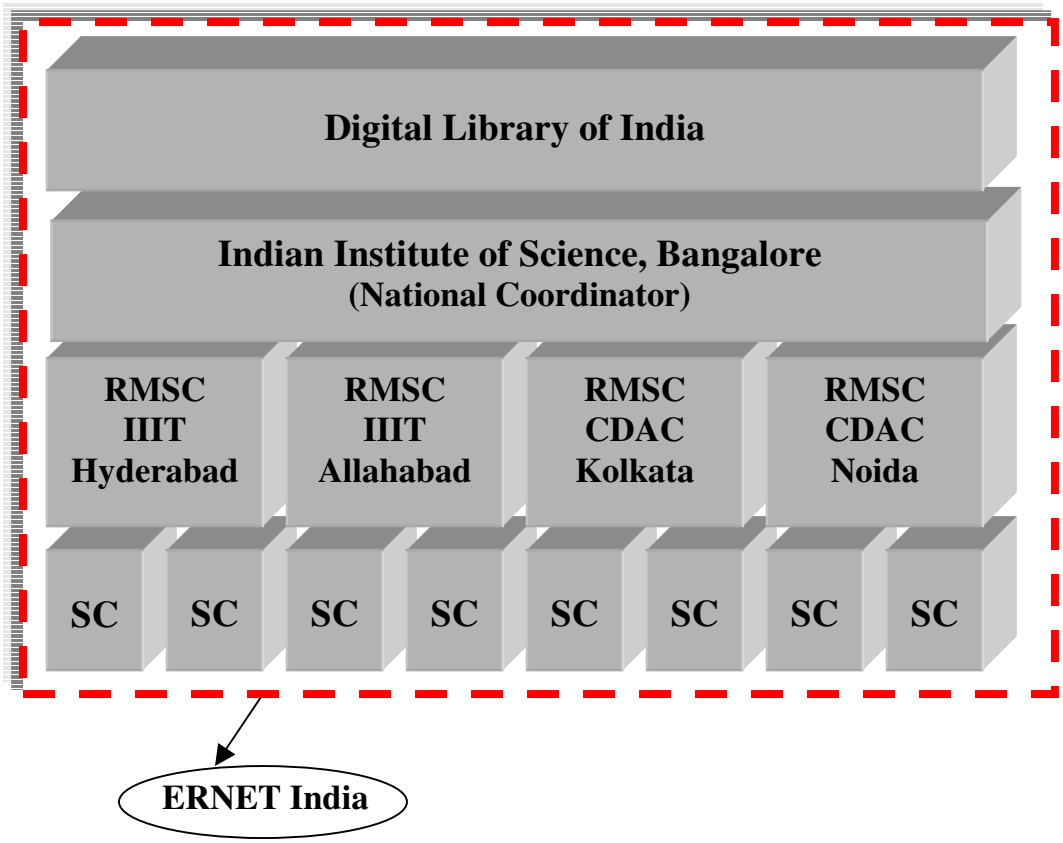


Fig. 2.12: ICT Infrastructure Provided by ERNET India

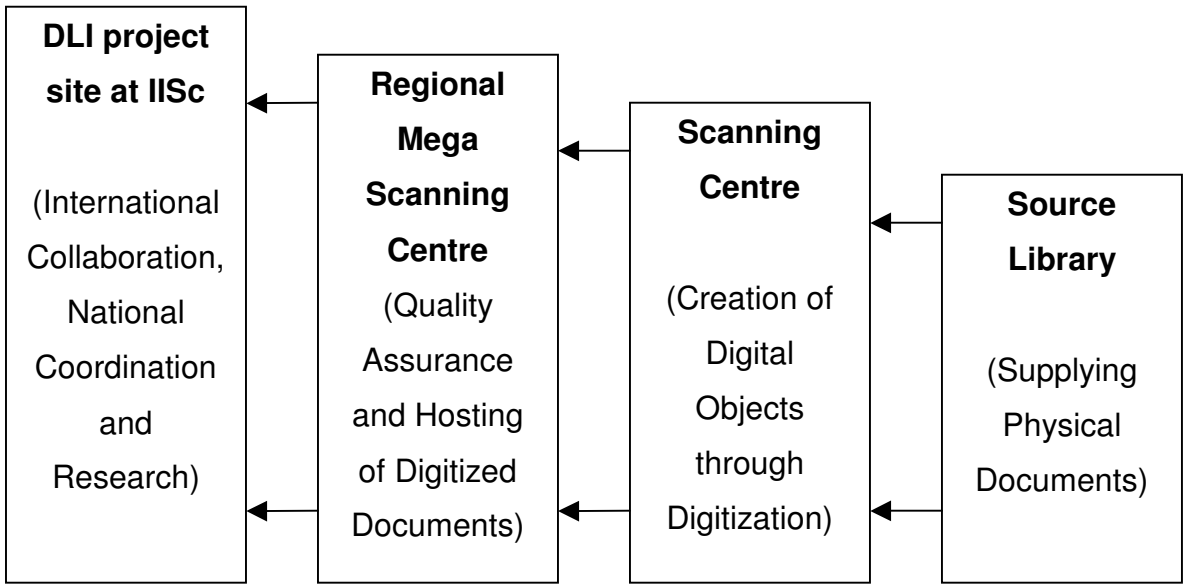


Fig. 2.13: Data Flow Diagram within DLI Project Partners

2.3 Spreading of DLI Project Sites Across the Country

Digital Library of India project sites spread over India through an array of scanning centres and source libraries, coordinated by four regional mega scanning centres and a national coordinator. Figure 2.14 illustrates nation-wide spread of DLI project sites. This is only an indicative map as on 17 May 2007. Few other DLI partners were added after that. This map indicates that DLI project sites are located in many states of India, particularly where resourceful libraries and institutions are located. DLI project partners also ensure that geographical representation should be reflected during the book selection process. For example, scanning centres in West Bengal ensure digitization of rare books of Bengali literature and culture.

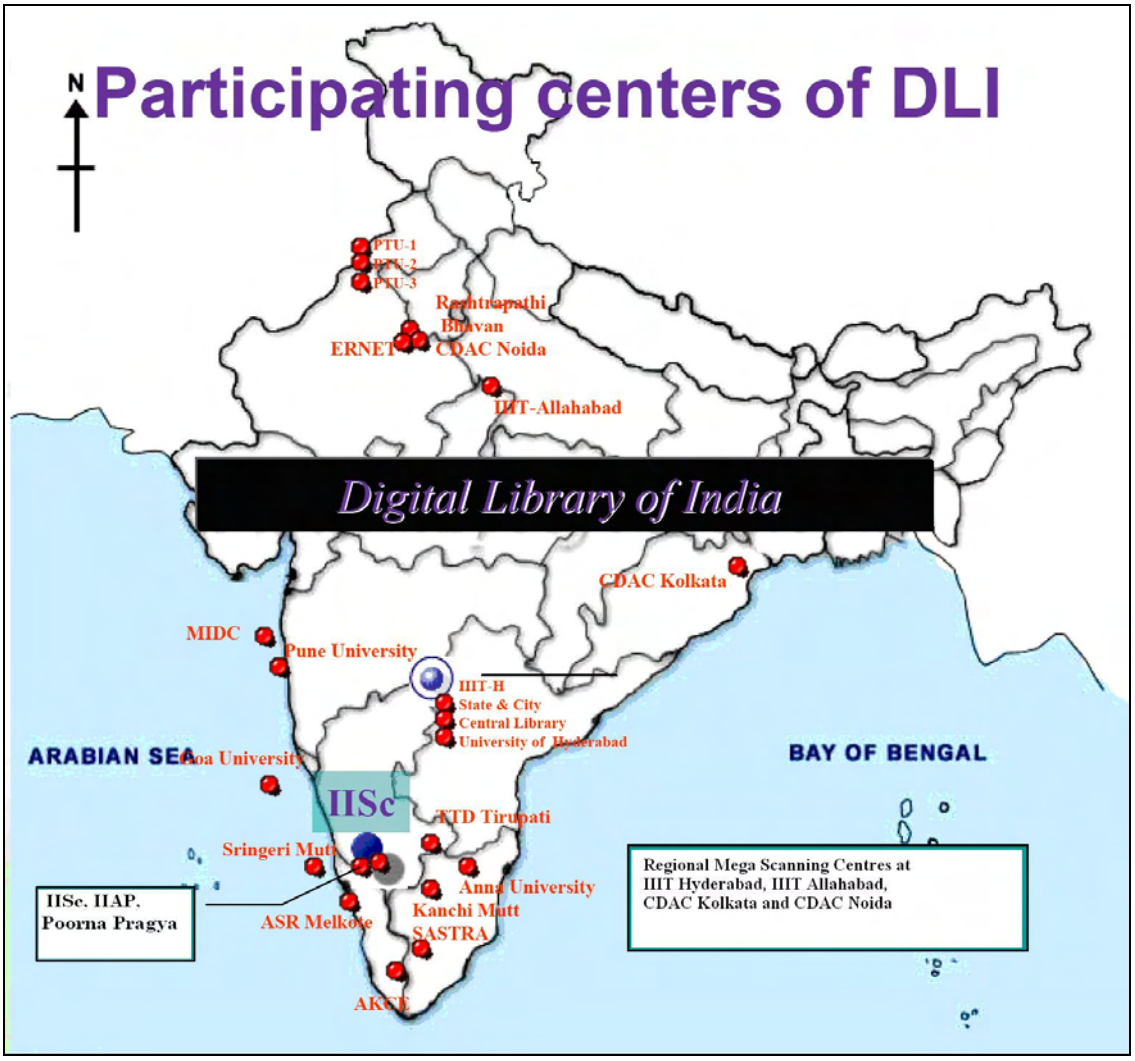


Fig. 2.14: Participating Centres of DLI across the Country

2.3.1 Regional Mega Scanning Centre – Kolkata

The Centre for Development of Advanced Computing (CDAC) Kolkata hosts Regional Mega Scanning Centre for the eastern zone of India. This Regional Mega Scanning Centre has a centralized digitization facility where some source libraries sent their books for digitization. This mega scanning centre also coordinates digitization activities in different scanning centres across the region. DLI portal in IISC Bangalore hosts digitized books from the following scanning centres and source libraries, operated under Regional Mega Scanning Centre in CDAC Kolkata:

- Bangiya Sahitya Parishad, Kolkata, West Bengal
- Bureau of Applied Economics and Statistics, Kolkata, West Bengal
- Central Library, Visva Bharati, Santiniketan, West Bengal
- Dr. Meghnad Saha Collection, Kolkata, West Bengal
- Hooghly Mohsin College, Hooghly, West Bengal
- Natya Sodh Sansthan (an Archive of Indian Theatre), Kolkata, West Bengal
- North Bengal State Library, Cooch Behar, West Bengal
- Raj Bhavan Library, Kolkata, West Bengal
- Ramakrishna Mission Institute of Culture, Kolkata, West Bengal
- State Central Library, Kolkata, West Bengal
- Tea Research Association Library (TRA), Jorhat, Assam
- The Asiatic Society, Kolkata, West Bengal
- Uttarpara Jayakrishna Public Library, Hooghly, West Bengal
- West Bengal Legislative Assembly House, Kolkata, West Bengal

2.3.2 Regional Mega Scanning Centre – Hyderabad

In Hyderabad, International Institute of Information Technology (IIIT) Hyderabad hosts Regional Mega Scanning Centre for the southern and central zones of India. This mega scanning centre has a centralized digitization facility where some source libraries sent their books for digitization. This mega scanning centre also coordinates digitization activities in different scanning centres across the

region. DLI portal at IIIT Hyderabad hosts digitized books from the following scanning centres and source libraries, operated under regional mega scanning centre in IIIT Hyderabad:

- Andhra Pradesh State Archives, Hyderabad, Andhra Pradesh
- City Central Library, Hyderabad, Andhra Pradesh
- Dr B R Ambedkar Open University (BRAOU), Hyderabad, Andhra Pradesh
- Ekalavya Publications, Bhopal, Madhya Pradesh
- Osmania University, Hyderabad, Andhra Pradesh
- Potti Sreeramulu Telugu University (PSTU) Library, Hyderabad, Andhra Pradesh
- Paropakarini Sabha, Andhra Pradesh
- Salar Jang Museum and Library, Hyderabad, Andhra Pradesh
- State Central Library, Hyderabad, Andhra Pradesh
- Tirumala Tirupati Devasthanams, Sri Venkateswara Digital Library, Tirupati, Andhra Pradesh
- University of Hyderabad, Andhra Pradesh

Sri Venkateswara Digital Library of Tirumala Tirupati Devasthanams, having a large scale scanning facility, is digitizing books from more than ten source libraries in the cities of Tirupati and Rajahmundry in Andhra Pradesh. Regional Mega Scanning Centre at IIIT Hyderabad has digitized books from some overseas partner of Universal Digital Library project. This activity can be referred as outsourced scanning activity, where a library in developed country is outsourcing its digitization activity to a scanning centre at a developing country. Some of the partners are:

- Food and Agriculture Organization of the United Nations (FAO), Italy (119 books digitized and hosted)
- Kansas City Public Library, United States of America (5396 books digitized and hosted)
- University of Washington Libraries, United States of America (1181 books digitized and hosted)

2.3.3 Regional Mega Scanning Centre – Noida

In Noida, Centre for Development of Advanced Computing (CDAC) Noida hosts Regional Mega Scanning Centre for the northern zone of India. This mega scanning centre has a centralized digitization facility where some source libraries sent their books for digitization. This centre also coordinates digitization activities in different scanning centres across the region. DLI portal at CDAC Noida hosts digitized books from the following scanning centres and source libraries, operated under regional mega scanning centre in CDAC Noida:

- Arya Samaj, New Delhi
- Association of Indian Universities, New Delhi
- Bharatiya Jnanpith, New Delhi
- Birla Institute of Technology and Science (BITS), Pilani, Rajasthan
- Delhi Engineering College, New Delhi
- Dyal Singh Public Library, New Delhi
- G.B. Pant University of Agriculture & Technology, Pantnagar, Uttarakhand
- Gurukul Kangari University, Haridwar, Uttarakhand
- Indian Agricultural Research Institute (IARI), New Delhi
- Indian Council for Cultural Relations (ICCR), New Delhi
- Kumaun University, Nainital, Uttarakhand
- Lal Bahadur Shastri National Academy of Administration, Musoorie, Uttarakhand
- Nagari Pracharini Sabha, Varanasi, Uttar Pradesh
- National Council of Educational Research and Training (NCERT), New Delhi
- Prajapita Brahma Kumaris Ishwariya Vishwa Vidyalaya, Abu Road, Rajasthan
- The Energy and Resources Institute (TERI), New Delhi
- University of Delhi, Delhi

2.3.4 Regional Mega Scanning Centre – Allahabad

Indian Institute of Information Technology Allahabad (IIITA) hosts Regional Mega Scanning Centre for the Uttar Pradesh state. This Centre has a centralized digitization facility where some source libraries send their books for digitization. This mega scanning centre also coordinates digitization activities in different scanning centres across this region. DLI portal at IISc Bangalore hosts digitized books from the following source libraries, operated under regional mega scanning centre in IIIT Allahabad:

- National Academy of Sciences, Allahabad
- Hindustani Academy, Allahabad
- Vijnan Parishad, Allahabad
- Ganga Nath Jha Sanskrit Vidyapeeth, Allahabad
- Prayag Sangeet Samiti, Allahabad
- Allahabad High Court, Allahabad
- Indian Institute of Technology Kanpur (IITK)
- Allahabad University, Allahabad

IIIT Allahabad has tied up with few other zonal public institutions for the second phase of DLI project. These institutions are:

- J.K. Institute of Applied Physics & Technology, Allahabad University
- Allahabad Public Library, Allahabad
- Hindi Sahitya Sammelan, Allahabad
- Allahabad Museum, Allahabad
- Govind Ballabh Pant Social Science Institute, Jhusi, Allahabad
- Sampurnanand Sanskrit Vishwavidyalaya, Varanasi
- Banaras Hindu University, Varanasi
- Mahatma Gandhi Kashi Vidyapeeth, Varanasi
- Central Institute of Higher Tibetian Studies, Sarnath, Varanasi
- Jagadguru Rambhadracharya Handicapped University, Chitrakoot Dham

2.3.5 Other Scanning Centres directly Operated under DLI National Coordinator

There are several other scanning centres associated with DLI project, which usually have no formal relation with any of the regional mega scanning centres. They usually get direction directly from the DLI national coordination office in IISc Bangalore and reporting. These scanning centres mainly digitize their own resources, if their collections are very rich. When a scanning centre completes digitization of its own resources, it may opt for digitizing collections from partnered source libraries. Some of these scanning centres also secured MCIT (Indian Ministry of Communication and Information Technology) grants for setting up the scanning facilities, and/or secured high performance book scanners from National Science Foundation of the United States, a sponsor of UDL project. The names of scanning centres, operated directly under DLI national coordinator, are given below:

- Academy of Sanskrit Research , Melkote, Karnataka
- Anna University, Chennai, Tamil Nadu
- Arulmigu Kalasligam College of Engineering (AKCE), Virudhunagar, Tamil Nadu
- Goa University, Goa
- Indian Institute of Astrophysics (IIAP), Bangalore, Karnataka
- Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya (Kanchi University), Kanchi, Tamil Nadu
- Maharashtra Industrial Development Corporation (MIDC), Mumbai, Maharashtra
- Poornapragna Vidyapeetha, Bangalore, Karnataka
- Punjab Technical University, Punjab
- Rashtrapati Bhavan, New Delhi
- Shanmugha Art, Science, Technology and Research Academy (SASTRA), Tanjavur, Tamil Nadu
- Sri Sringeri Sharada Peetham (Sringeri Mutt), Sringeri, Karnataka
- University of Pune, Pune, Maharashtra

Table 2.1 presents a comprehensive list of DLI partners across India, including funding agencies, regional mega scanning centres, scanning centres, source libraries and infrastructure partner. This Table also indicates a huge involvement of ninety plus organizations to make this an initiative of national importance.

Table 2.1: Comprehensive List of DLI Project Partners in India

Name of DLI Partner	State	Role	Line Authority
Indian Institute of Science, Bangalore [www.iisc.ernet.in]	Karnataka	National Coordinator	UDL Carnegie Melon University
National Digital Libraries Cell, Ministry of Communication and Information Technology (MCIT) [www.mit.gov.in/]	Delhi	Funding, Monitoring & Evaluation	Government of India
Office of the Principal Advisor to the Government of India [www.psa.gov.in/]	Delhi	Funding	Government of India
Technology Development for Indian Languages (TDIL), Ministry of Communication and Information Technology (MCIT) [http://tdil.mit.gov.in/]	Delhi	Research Partner	Government of India
ERNET India [www.eis.ernet.in/]	Delhi	ICT Infrastructure Provider	DLI, IISc
International Institute of Information Technology, Hyderabad (IIITH) [www.iiit.net/ & www.iiith.ac.in]	Andhra Pradesh	RMSC	DLI, IISc
Indian Institute of Information Technology, Allahabad (IIITA) [www.iiita.ac.in]	Uttar Pradesh	RMSC	DLI, IISc
			Contd…

Name of DLI Partner	State	Role	Line Authority
Centre for Development of Advanced Computing (CDAC), Kolkata [www.kolkatacdac.in]	West Bengal	RMSC	DLI, IISc
Centre for Development of Advanced Computing (CDAC) Noida [http://cdacnoida.in/]	Uttar Pradesh	RMSC	DLI, IISc
Academy of Sanskrit Research, Melkote [www.sanskritacademy.org/]	Karnataka	Scanning Centre	DLI, IISc
Anna University, Chennai [www.annauniv.edu]	Tamil Nadu	Scanning Centre	DLI, IISc
Arulmigu Kalasaligam College of Engineering (AKCE), Virudhunagar [www.kalasalingam.ac.in/]	Tamil Nadu	Scanning Centre	DLI, IISc
Goa University [www.unigoa.ac.in/]	Goa	Scanning Centre	DLI, IISc
Indian Institute of Astrophysics, Bangalore [www.iiap.res.in/]	Karnataka	Scanning Centre	DLI, IISc
Indian National Science Academy, New Delhi [www.insaindia.org/, www.insa.ac.in/]	Delhi	Content Partner	DLI, IISc
Maharashtra Industrial Development Corporation (MIDC), Mumbai [www.midcindia.org/]	Maharashtra	Scanning Centre	DLI, IISc
Poornapragna Vidyapeetha, Bangalore	Karnataka	Scanning Centre	DLI, IISc
Punjab Technical University, Punjab [www.ptu.ac.in/]	Punjab	Scanning Centre	DLI, IISc
Rashtrapati Bhavan, New Delhi [www.presidentofindia.nic.in/]	Delhi	Scanning Centre	DLI, IISc
			Contd…

Name of DLI Partner	State	Role	Line Authority
Shanmugha Art, Science, Technology & Research Academy (SASTRA), Tanjavur [www.sastra.edu/]	Tamil Nadu	Scanning Centre	DLI, IISc
Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya (Kanchi University), Kanchi [www.kanchiuniv.ac.in/]	Tamil Nadu	Scanning Centre	DLI, IISc
Sri Sringeri Sharada Peetham, Sringeri [www.sringerisharadapeetham.org]	Karnataka	Scanning Centre	DLI, IISc
University of Pune [www.unipune.ernet.in/]	Maharashtra	Scanning Centre	DLI, IISc
Allahabad High Court [www.allahabadhighcourt.in/]	Uttar Pradesh	Source Library	RMSC Allahabad
Andhra Pradesh State Archives and Research Institute, Hyderabad [www.apts.gov.in/andhrastatearchives/]	Andhra Pradesh	Source Library	RMSC Hyderabad
Besant Theosophical College, Madanapalli [www.btcollege.org]	Andhra Pradesh	Source Library	RMSC Hyderabad
C.P. Brown Library, Kadapa	Andhra Pradesh	Source Library	RMSC Hyderabad
City Central Library, Hyderabad	Andhra Pradesh	Scanning Centre	RMSC Hyderabad
Dr. B. R. Ambedkar Open University (BRAOU), Hyderabad [www.braou.ac.in/]	Andhra Pradesh	Scanning Centre	RMSC Hyderabad
			Contd…

Name of DLI Partner	State	Role	Line Authority
Eklavya Publications, Bhopal [http://eklavya.in/]	Madhya Pradesh	Source Library	RMSC Hyderabad
Ganga Nath Jha Sanskrit Vidyapeeth, Allahabad	Uttar Pradesh	Source Library	RMSC Allahabad
Hindi Sahitya Sammelan, Allahabad	Uttar Pradesh	Source Library	RMSC Allahabad
Hindustani Academy, Allahabad	Uttar Pradesh	Source Library	RMSC Allahabad
Indian Institute of Technology, Kanpur [www.iitk.ac.in/]	Uttar Pradesh	Source Library	RMSC Allahabad
National Academy of Sciences, Allahabad [www.nasi.org.in/]	Uttar Pradesh	Source Library	RMSC Allahabad
Oriental Research Institute, Tirupati	Andhra Pradesh	Source Library	RMSC Hyderabad
Osmania University, Hyderabad [www.osmania.ac.in/]	Andhra Pradesh	Scanning Centre	RMSC Hyderabad
Par Informatics Ltd (PAR), Hyderabad	Andhra Pradesh	Vendor	RMSC Hyderabad
Paropakarini Sabha [http://paropkarini.org]	Andhra Pradesh	Source Library	RMSC Hyderabad
PCR College, Kakinada	Andhra Pradesh	Source Library	RMSC Hyderabad
Potti Sreeramulu Telugu University (PSTU), Hyderabad [www.teluguuniversity.ac.in/]	Andhra Pradesh	Scanning Centre	RMSC Hyderabad
Prayag Sangeet Samiti, Allahabad [www.prayagsangeetsamiti.in/]	Uttar Pradesh	Source Library	RMSC Allahabad

Contd…

Name of DLI Partner	State	Role	Line Authority
Salar Jang Museum and Library, Hyderabad [www.salarjungmuseum.in/]	Andhra Pradesh	Scanning Centre	RMSC Hyderabad
Saraswata Niketanam Library, Vetapalem [http://saraswataniketanam.org/]	Andhra Pradesh	Source Library	RMSC Hyderabad
Sri Padmavathi Mahila University, Tirupati [www.padmavatiwomen-univ.org/]	Andhra Pradesh	Source Library	RMSC Hyderabad
Sri Venkateswara Digital Library, Rajahmundry	Andhra Pradesh	Source Centre	RMSC Hyderabad
Sri Venkateswara Infosys (SVI), Tirumala Tirupati Devasthanams, Tirupati [www.tirtumala.org/]	Andhra Pradesh	Vendor	RMSC Hyderabad
Sri Venkateswara University, Tirupati [www.svuniversity.in/]	Andhra Pradesh	Source Library	RMSC Hyderabad
State Central Library, Hyderabad	Andhra Pradesh	Scanning Centre	RMSC Hyderabad
The Kuppuswami Sastri Reserach Institute, Mylapore	Tamil Nadu	Source Library	RMSC Hyderabad
Thrinaina Informatics Ltd. (TIL), Hyderabad [www.thrinaina.com/]	Andhra Pradesh	Vendor	RMSC Hyderabad
Sri Venkateswara Digital Library, Tirumala Tirupati Devasthanams, Tirupati [www.tirumala.org/]	Andhra Pradesh	Scanning Centre	RMSC Hyderabad
University of Allahabad, Allahabad [www.allduniv.ac.in/]	Uttar Pradesh	Source Library	RMSC Allahabad
University of Hyderabad [www.uohyd.ernet.in/]	Andhra Pradesh	Scanning Centre	RMSC Hyderabad
			Contd…

Name of DLI Partner	State	Role	Line Authority
Vardhamana Samajam, Nellore	Andhra Pradesh	Source Library	RMSC Hyderabad
Victoria Library, Machalipatnam	Andhra Pradesh	Source Library	RMSC Hyderabad
Vijnan Parishad, Allahabad	Uttar Pradesh	Source Library	RMSC Allahabad
Vivekananda College, Hyderabad	Andhra Pradesh	Source Library	RMSC Hyderabad
Bangiya Sahitya Parishad, Kolkata	West Bengal	Source Library	RMSC Kolkata
Bureau of Applied Economics and Statistics, Kolkata	West Bengal	Source Library	RMSC Kolkata
Dr. Meghnad Saha Collection, Kolkata [www.saha.ac.in/]	West Bengal	Source Library	RMSC Kolkata
Hooghly Mohsin College, Hooghly [www.hooghlymohsincollege.org/]	West Bengal	Source Library	RMSC Kolkata
Natya Sodh Sansthan, Kolkata [http://natyashodh.org/]	West Bengal	Source Library	RMSC Kolkata
North Bengal State Library, Cooch Behar	West Bengal	Source Library	RMSC Kolkata
Raj Bhavan Library, Kolkata [http://rajbhavankolkata.gov.in; http://rajbhavankolkata.nic.in]	West Bengal	Source Library	RMSC Kolkata
Ramakrishna Mission Institute of Culture, Kolkata [www.sriramakrishna.org/]	West Bengal	Source Library	RMSC Kolkata
State Central Library, Kolkata	West Bengal	Source Library	RMSC Kolkata
Tea Research Association (TRA), Jorhat [http://jorhat.nic.in/ttc.htm]	Assam	Source Library	RMSC Kolkata
			Contd…

Name of DLI Partner	State	Role	Line Authority
The Asiatic Society, Kolkata [www.asiaticsocietycal.com/]	West Bengal	Source Library	RMSC Kolkata
Uttarpara Jayakrishna Public Library Hooghly	West Bengal	Scanning Centre	RMSC Kolkata
Visva Bharati, Santiniketan [www.visva-bharati.ac.in/]	West Bengal	Scanning Centre	RMSC Kolkata
West Bengal Legislative Assembly House, Kolkata	West Bengal	Source Library	RMSC Kolkata
Arya Samaj New Delhi [www.aryasamaj.com/]	Delhi	Source Library	RMSC Noida
Association of Indian Universities, New Delhi [www.aiuweb.org/]	Delhi	Source Library	RMSC Noida
Bharatiya Jnanpith, New Delhi [http://jnanpith.net/]	Delhi	Scanning Centre	RMSC Noida
Birla Institute of Technology and Science (BITS), Pilani [www.bits-pilani.ac.in/]	Rajasthan	Source Library	RMSC Noida
Delhi College of Engineering, Delhi [www.dce.edu/]	Delhi	Source Library	RMSC Noida
Dyal Singh Public Library, New Delhi	Delhi	Source Library	RMSC Noida
Govind Ballabh Pant University of Agriculture and Technology, Pantnagar [www.gbpuat.ac.in/]	Uttarakhand	Source Library	RMSC Noida
Gurukul Kangri University, Haridwar [www.gkvharidwar.org/]	Uttarakhand	Scanning Centre	RMSC Noida
Indian Agricultural Research Institute (IARI), New Delhi [www.iari.res.in/]	Delhi	Source Library	RMSC Noida
			Contd…

Name of DLI Partner	State	Role	Line Authority
Indian Council for Cultural Relations (ICCR), New Delhi [www.iccrindia.org/]	Delhi	Source Library	RMSC Noida
Kumaun University, Nainital [www.kumaununiversity.org/]	Uttarakhand	Source Library	RMSC Noida
Lal Bahadur Shastri National Academy of Administration, Mussoorie [www.lbsnaa.ernet.in/]	Uttarakhand	Source Library	RMSC Noida
Nagari Pracharini Sabha, Varanasi [http://tempweb34.nic.in/xnagari/]	Uttar Pradesh	Scanning Centre	RMSC Noida
National Council of Educational Research and Training (NCERT), New Delhi [http://ncert.nic.in/]	Delhi	Source Library	RMSC Noida
Prajapita Brahma Kumaris Ishwariya Vishwa Vidyalaya, Abu Road [www.brahmakumaris.com/]	Rajasthan	Source Library	RMSC Noida
The Energy and Resources Institute (TERI), New Delhi [www.teriin.org/]	Delhi	Source Library	RMSC Noida
University of Delhi [www.du.ac.in/]	Delhi	Source Library	RMSC Noida

2.4 Installation of Book Scanners across States of India

As on 17 May 2007, Regional Mega Scanning Centres and Scanning Centres across India installed about 171 high performance book scanners, particularly book scanners with Minolta and Zeutschel brands. Initially, some book scanners were procured utilizing financial assistance from the National Science Foundation. Later, more book scanners were procured utilizing financial assistance of Government of India. Figure 2.15 indicates availability of high capacity book scanners among regional mega scanning centres and other scanning centres. This Figure also indicates that IIIT Hyderabad possesses maximum number of book scanners, 45 out of 171, i.e., more than one fourth of

total number of scanners. Obviously, this RMSC has huge digitization capacity, unparallel to any other RMSCs. This Figure also depicts that amongst two brands of book scanners, Zeutschel-branded scanners outnumbered Minolta scanners, i.e., 100 out of 171. Figure 2.16 provides information on availability of book scanners across Indian states, where Andhra Pradesh stands first (54.4%), Uttar Pradesh second (15.8%) and Tamil Nadu third (8.2%). Figure 2.16 provides a mapping of scanning partners across India, which shows many states of India are left behind in this project until now. This means, Digital Library of India portals may contain insignificant number of books, published in those non-participating states of India or in the states' official languages.

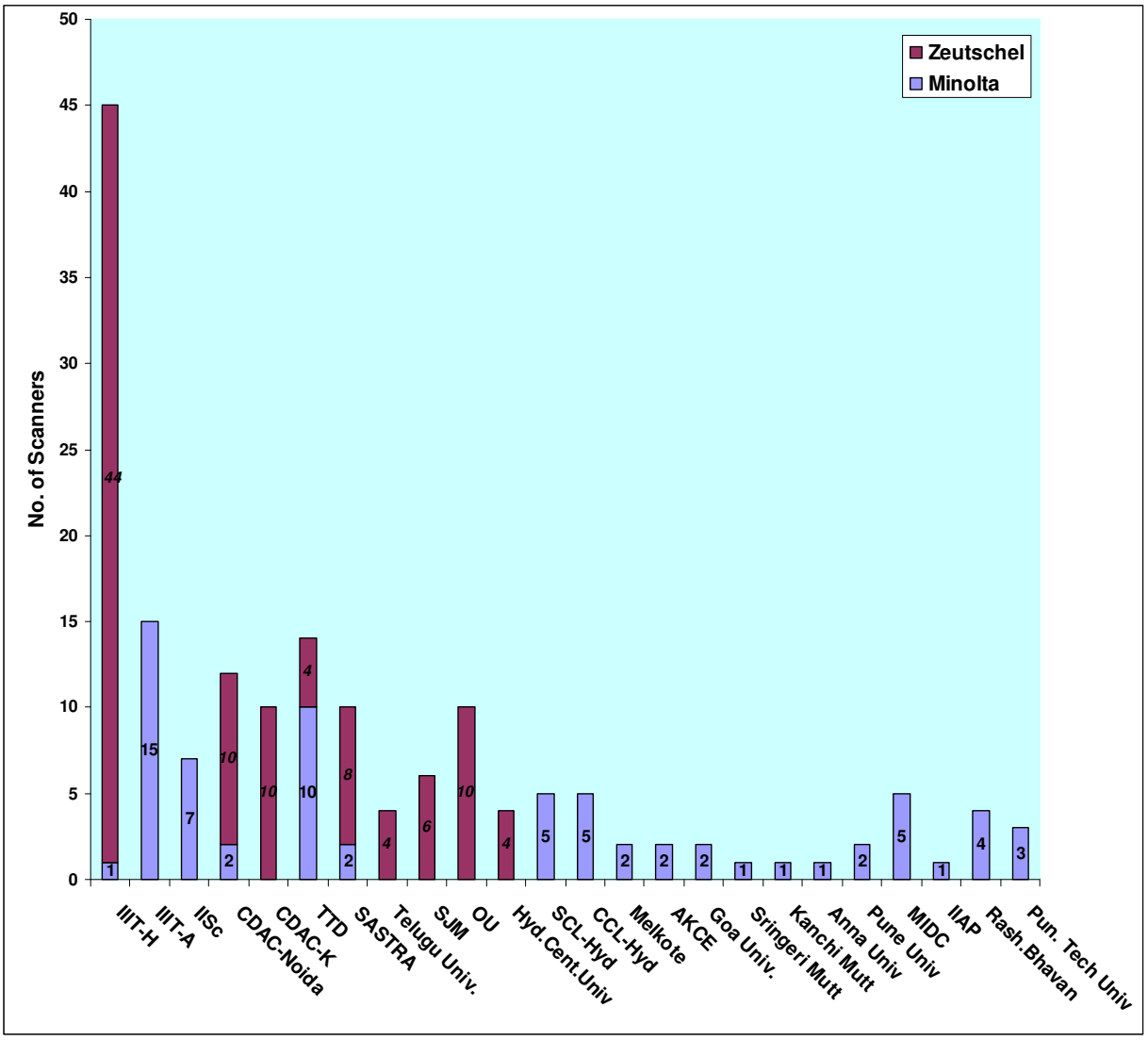


Fig. 2.15: Scanner Locations across India

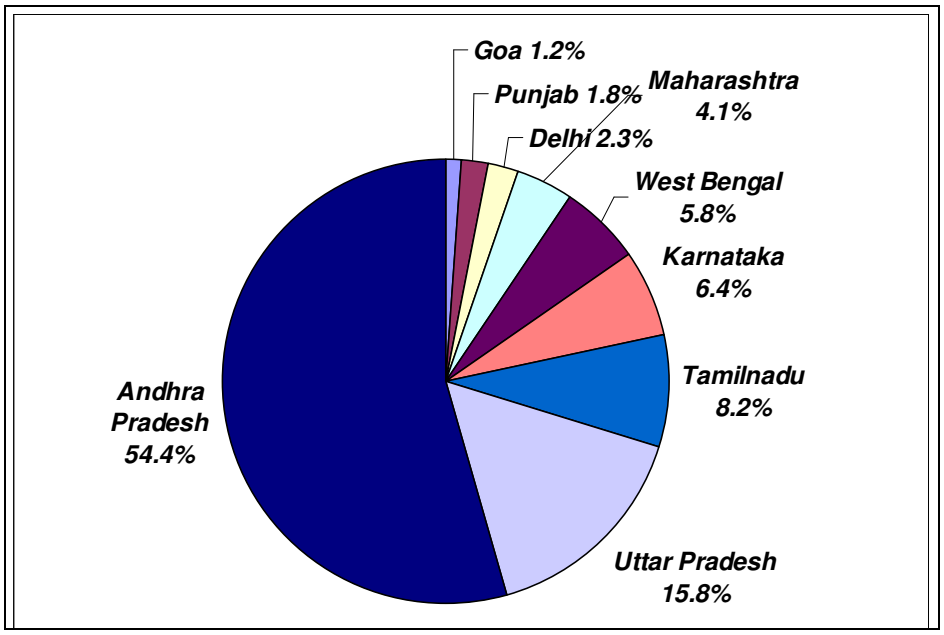


Fig. 2.16: State-wise Distribution of Book Scanners (As on 17 May 2007)

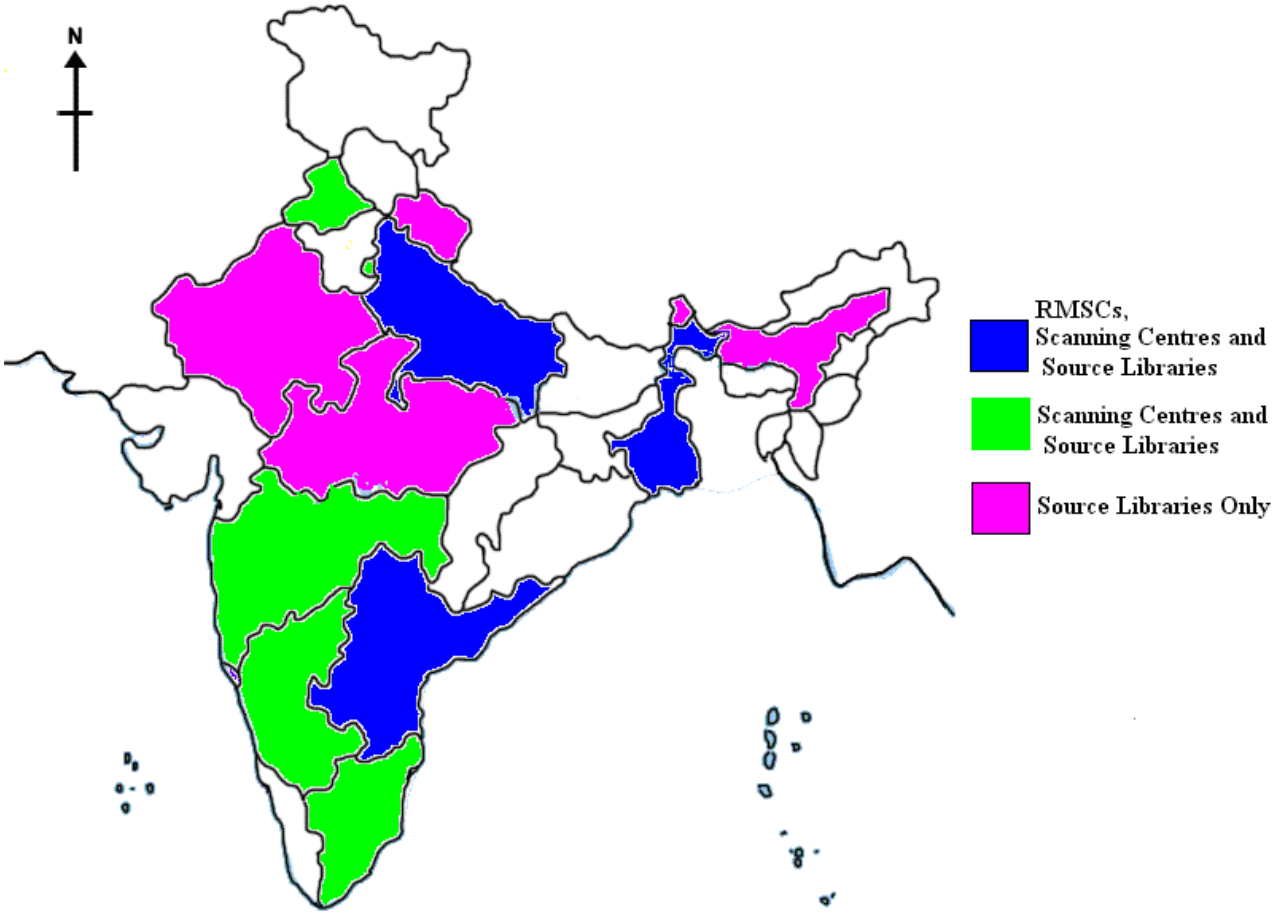


Fig. 2.17: Distribution of DLI Scanning Partners Across States of India

2.5 Funding Sources of DLI Project

Digital Library of India project is primarily funded by two national and one overseas agencies, namely Ministry of Communication and Information Technology (MCIT), Government of India; Office of the Principal Advisor to the Government of India; and the National Science Foundation, United States of America. Initially the Universal Digital Library project was funded by the National Science Foundation (NSF), a research funding agency in United States of America. NSF provided seed money to Indian UDL partners for the procurement of high performance book scanners and development of essential digital library software, when UDL project extended to India. Using NSF fund, first few book scanners were procured in India for the first few DLI scanning centres. The Office of the Principal Scientific Advisor to the Government of India has provided financial support to Indian Institute of Science, Bangalore for the establishment and maintenance of national project coordination centre there. Indian Ministry of Communication and Information Technology (MCIT) is funding many DLI-linked projects at various partner centres of the Digital Library of India, including regional mega scanning centres and local scanning centres. Since the initiation of DLI Project in 2002, MCIT has been generously providing financial supports to DLI project partners in different phases of DLI project cycle. Table 2.2 provides a comprehensive list of the DLI-linked projects supported by the MCIT. This Table also indicates that MCIT is the major contributor for development and maintenance of ICT infrastructure in DLI project sites across India. ICT infrastructure in DLI project includes terabyte Web servers for hosting DLI portals, broadband connectivity, local servers and workstations at the scanning centres and regional mega scanning centres. Figure 2.19 identifies that financial supports received by DLI-linked projects vis-à-vis other digitization projects in the country. Seventeen (85%) DLI-linked projects received MCIT financial assistance compared to three (15%) other digitization projects. This also signifies that National Digital Libraries Cell of MCIT stands very favorable to the Digital Library of India initiative, than any other digital library initiatives in the country.

The regional mega scanning centres and local scanning centres of DLI are free to generate their additional resources for the self-sustainability of the project by collaborating with the state authorities and non-government trusts including religious trusts and industrial agencies. Some scanning centres, in collaboration with private vendors, also undertake out-sourced digitization activities from the overseas clients as a spin-off of the project. DLI project also accepts donations from individuals and institutions. Donations are used for activities such as selection and acquisition of books, packing and shipping costs, and quality assurance of the digitized contents.

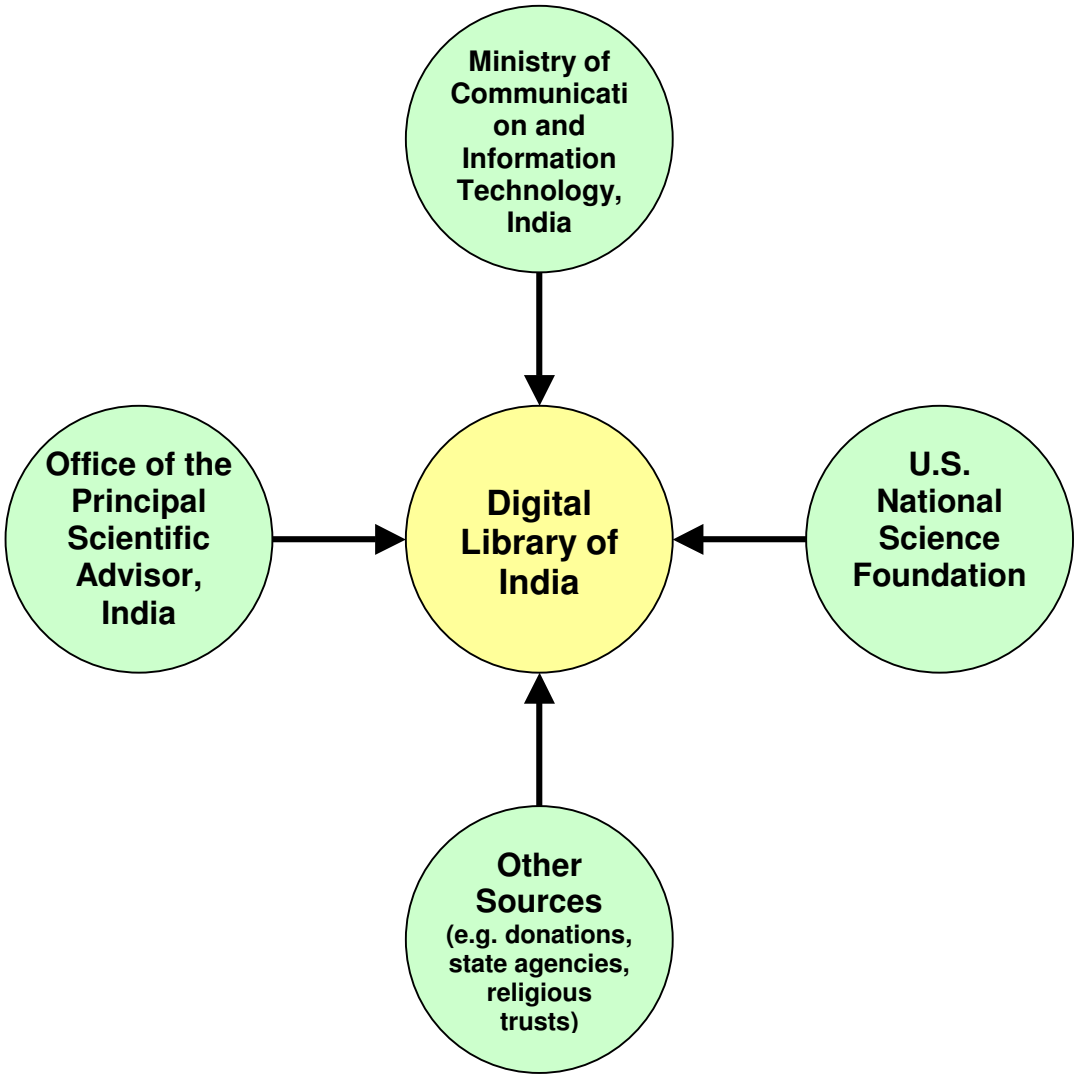


Fig. 2.18: Funding Sources of DLI Project

Table 2.2: DLI-linked Projects supported by MCIT

SL. No.	Name of Project and Implementing Agency	Status of the Project
1.	Establishment of Digital Library of India *By ERNET India*	Project ongoing. Installed servers at 13 Nodal centres. Provided the broadband connectivity to 6 nodal centres.
2.	Setting up of Scanning Centers in Uttar Pradesh for participation in the Universal Digital Library project *By IIIT Allahabad*	Project completed. The scanned data Web enabled
3.	Setting up of Scanning Centers in Maharashtra for participation in the Universal Digital Library project *By Maharashtra Industrial Development Corporation (MIDC), Mumbai*	Project completed. Scanned data Web enabled
4.	Setting up of Scanning Centers at Hyderabad for participation in the Universal Digital Library Project *By City Central Library and State Central Library, Hyderabad*	Project completed. Scanned data Web enabled
5.	Advaita Sarada Project (Digitization of Ancient Manuscripts in South Indian Languages pertaining to the Vedas, Vedangus, Upanishads and other Sastric Studies) *By Sri Sringeri Sharada Peetham, Sringeri, Karnataka*	Project completed. Scanned data Web enabled.
6.	Digital Archiving for Preservation of Rare Manuscripts and Old Magazines from mid-19th Century to 1960 available with Nagari Pracharini Sabha, Varanasi *By CDAC Noida*	Project completed. Scanned data Web enabled.

7.	Digital Archiving for Preservation of Rare Manuscripts and Folios Available with Namgyal Institute of Tibetology, Sikkim *By Namgyal Institute of Tibetology, Sikkim with technical support of CDAC Kolkata*	Ongoing
8.	Setting up of Scanning Center at Sringeri, Karnataka for participation in the Universal Digital Library project *By Sri Sringeri Sharada Peetham, Sringeri, Karnataka*	Project completed. Scanned data Web enabled.
9.	Setting up of Scanning Center at Goa for participation in the Universal Digital Library project *By University of Goa*	Project ongoing. Scanned data Web enabled.
10.	Setting up of Scanning Center at Hyderabad for participation in Universal Digital Library project *By University of Hyderabad*	Project ongoing. Scanned data Web enabled.
11.	Setting up of scanning centre at Bharatiya Jnanpith for participation in the Universal Digital Library project *By Bharatiya Jnanpith, New Delhi*	Project completed. Scanned data Web enabled.
12.	Setting up of Scanning Centres in Maharashtra for participations in the Universal Digital Library Project *By University of Pune*	Project completed. Scanned data Web enabled.
13.	Mobile Digital Library Project (Print Your Own Book) *By CDAC Noida*	Project completed. Books scanned, printed and distributed in schools.
14.	Creation of Digital Library of Books in President House *By CDAC Noida*	Project completed. Scanned data Web enabled.
15.	Establishing Centres for Digital Archiving and Creation of Rare Knowledge pertaining to Ayurvedic Medicine,	Project completed. Scanned data Web

	Integrating & Showcasing of the Content Created through Digital Library Outlet for Uttarakhand State Government *By CDAC Noida*	enabled.
16.	Digital Library of India - Phase 2: Creating a Large Collection of Wide National Interest *By IIIT Hyderabad*	Ongoing (Project initiated on 19th October 2007)
17.	Coordination, Web Hosting & Maintenance of Digital Library of India *By IISc Bangalore*	Ongoing (Project initiated on 30 October 2007)
Source: *National Digital Libraries Cell* www.mit.gov.in/default.aspx?id=325, as on 2 July 2008		

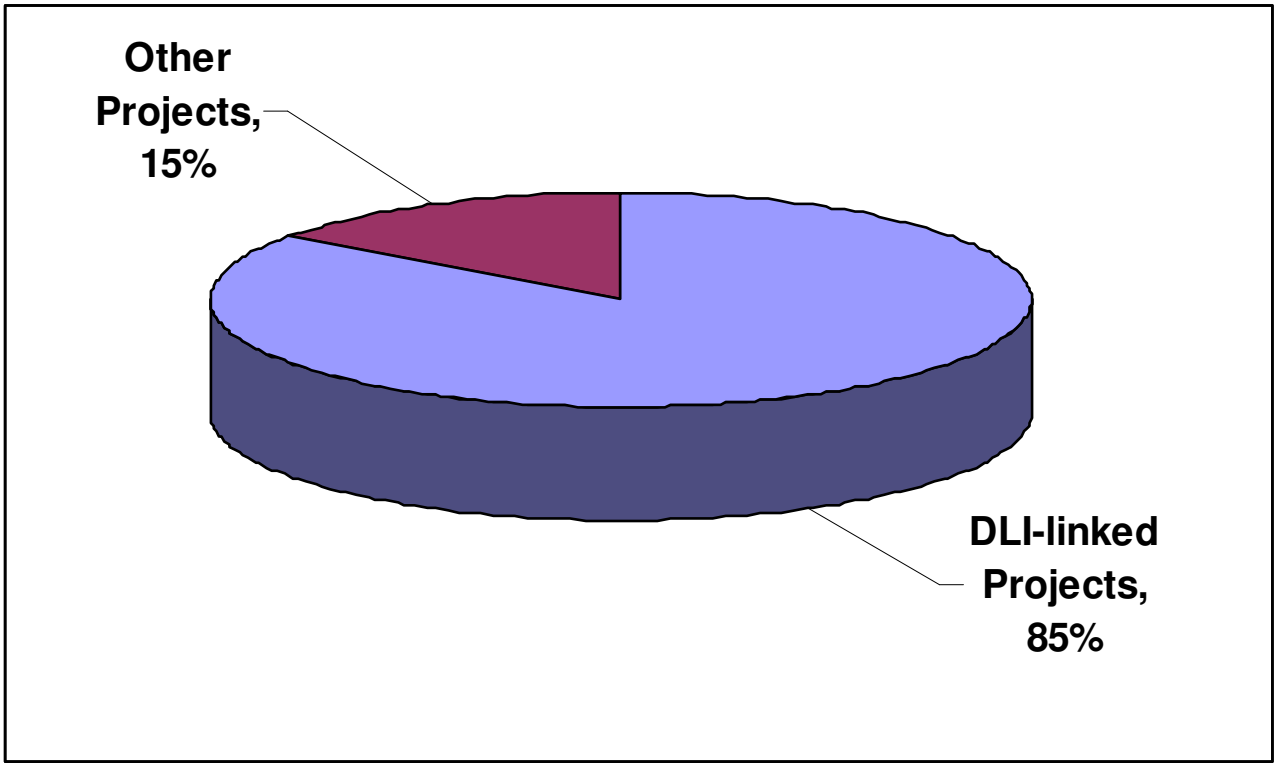


Fig. 2.19: MCIT-supported Digitization Projects

2.6 Partnership Pattern in DLI Project

Digital Library of India project is a flagship project in India that began with international collaboration. Over the time, this project built up successful multi-stake partnership, where different kinds of agencies joined hands to develop synergies in building up large scale digital library of the collective wisdom. The partners were drawn from academic institutions, cultural institutions, research agencies, religious institutions, government agencies and industrial agencies. Universities and colleges represent academic institutions, whereas museums, archives, public libraries represent cultural institutions. Religious institutions consist of religious bodies and trusts such as Tirumala Tirupati Devasthanam, and Sri Sringeri Sharada Peetham. The Centre for Development of Advanced Computing (CDAC), Academy of Sanskrit Research (ASR), etc. represent research agencies that became active partners of the DLI project. Many government agencies are also participating in this project. The Rashtrapati Bhavan (President's House), Raj Bhavan (Governor's House), West Bengal State Legislative Assembly are the examples of highly prestigious government offices taking part in this project. This project also expands partnership with corporate houses and industrial agencies such as Microsoft India, Thrinaina Informatics Limited and PAR Informatics Limited. DLI project successfully adopted public-private partnership (PPP) model for digitizing a large volume of literature. This project also helped in formation of a new industry segment to deal with outsourced digitization services, both from Indian and overseas organizations. Many enthusiastic entrepreneurs established companies to handle digitization activities at DLI scanning centres across India, as these newly established scanning centres were designed to be operated through third-party vendors for achieving operational efficiency. While professional librarians in the host institutions carefully evaluate metadata of each scanned document, data-entry operators and scanning assistants, provided by the third party vendors, are engaged in activities such as scanning, image editing and data entry. Thus, this project also helped in capacity building of private enterprises in India, who later on participated in many digitization projects across India, even overseas.

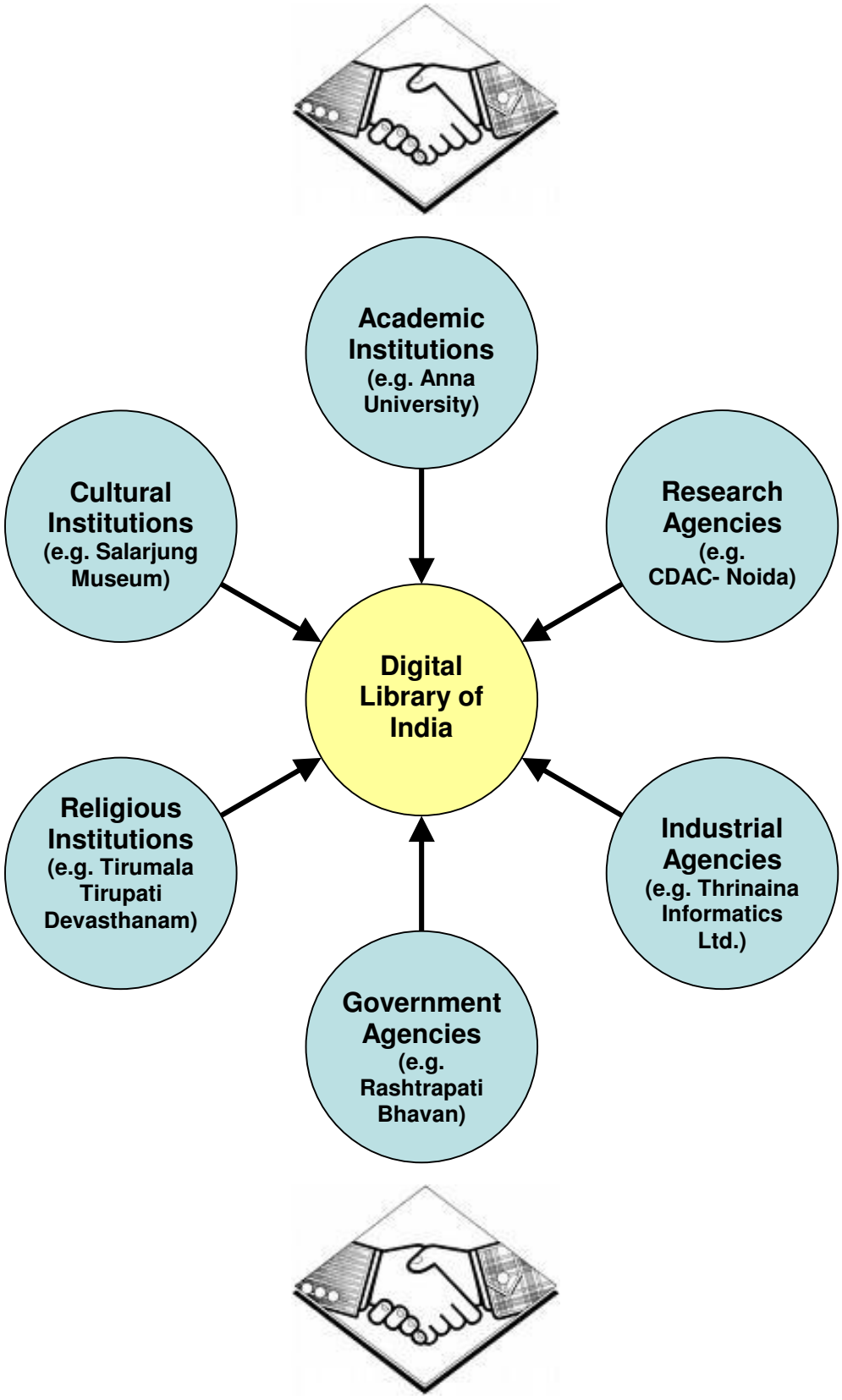


Fig. 2.20: Partnership Pattern in DLI Project

2.7 Outreaching DLI Portals at Remote Areas through Mobile Library

CDAC Noida has initiated a mission-mode project *Digital e-Library* (Dware Dware Gyan Sampadaa/ Providing Books at Your Doorsteps) in bringing the million books, available with the DLI and UDL portals, to the doorsteps of common citizens located at remote locations. Its Mission is “Internet enabled Mobile Digital Library brought to use of the common citizen for promoting literacy”. Different places such as schools in villages and other remote areas are covered under this programme to promote literacy and demonstrate the use of technology for masses, particularly in close proximity to the National Capital Region (NCR) and in particular states in northern and central India such as Delhi, Uttar Pradesh, Haryana, Madhya Pradesh, Himachal Pradesh, Uttarakhand, Punjab, Rajasthan, and Bihar. The schools, adult learning centres, public libraries or local panchayats can invite this mobile digital library for accessing resources and services attached with the mobile unit. This initiative makes use of a mobile van with satellite connection for Internet connectivity. The van is fitted with necessary accessories for providing bound books to the end user from a single point. The mobile van is equipped with:

- Dish Antenna for Internet Connectivity
- Multimedia Laptop
- Laser Printer
- Scorer and Cutter
- Automatic Book Binder

Other than books available in DLI portals, some full-text e-books are also made available locally in its portal in either one of DJVU, PDF and DOC formats. Some of the books are meant for neo-literates in spreading functional literary and lifelong learning in remote areas. The locally available e-books are written in Hindi language, as this initiative is covered in Hindi-speaking states only.

Fig. 2.21: Portal of Digital e-Library – An Outreach Initiative for DLI

2.8 Processes of Digitization Followed in the DLI Project

Digital Library of India, as a country partner of the Universal Digital Library initiative, follows the same standards and processes as adopted across the UDL project partners in order to achieve interoperability of all UDL portals and Web-servers. The metadata structure of this project as well as the digitization processes are well defined and well documented, so that every scanning partner can easily follow them (Michalek, 2003).

DLI project has well-defined process workflow, where process workflow is divided into three major elements, viz., pre-scanning process, scanning process and post-scanning process. Each process element also has a number of pre-defined activities. Figure 2.22 depicts the DLI process workflow, where input is a physical document and outcome is a digital document with essential metadata information. In between the input and outcome of the digitization process, rigorous procedures of standard compliance and quality assurance are applied to make the digital surrogates highly compatible for providing digital library service.

In the pre-scanning process, activities include selection of books, procurement of selected books from the source library, checking duplication of work (whether already digitized), generation of barcodes for non-duplicate books and then barcode assigned books issued to contractor for scanning. The whole pre-scanning process is shown in Figure 2.23.

The source libraries usually send a formal request o DLI authority through a Request for Participation Form (Annex 2), followed by signing up a memorandum of understanding (MoU), a legal agreement, with DLI project to establish partnership between DLI project and individual libraries. This agreement is established to safeguard the joint ownership of digitized documents. Books selected for digitization have importance in socio-cultural and literacy contexts of the country in general and state in particular. Sometimes the documents are selected for scanning, published by the parent institution of source library. For example, when the Asiatic Society, Kolkata becomes a source library, importance should be given to books published by the same organization, but

presently out-of-print or books not in circulation (not available in normal trade channel) are considered.

Proper care is taken when books from source libraries are sent to scanning centres. Most of the selected books are either very rare, or out of circulation in normal trade channel, or very brittle. Some books may not be available elsewhere within the country. Thus, damage and loss of books are very sensible issues to the both partners – source library and scanning centre.

When books are sent to a scanning centre, books are counted and information on sent books is made available with the library authority. Copyright date is verified to ensure that copyright policy of DLI is properly maintained. After completion of scanning or even in the case of rejection, books are properly marked and sent back to source library. The book movement within the source library and scanning centre is graphically illustrated in Figure 2.24. This Figure also provides a detail of pre-scanning and scanning processes followed in scanning centre. This flow chart is prepared by RMSC Kolkata to train their scanning assistants and data entry operators. The book movement within a source library is a continuous process, after completion of one batch of books another batch follows.

After the pre-scanning process, the scanning process is undertaken in a scanning centre by digitization contractor with the help of scanning operator and metadata-entry operator or metadata librarian. Several activities being undertaken during scanning process as illustrated in Figure 2.25 are described below:

- The books are scanned by an scanning operator at a particular location in the scanning centre with an allotted book scanner;
- The operator creates the structural meta information for the book he / she scans;
- The operator/ meta data entry operator enters the administrative (admin) meta information for the books scanned on a particular day;

- The scanned books undergo image processing and OCR and also quality control is done by the contractors/supervisors;
- The scanned and processed books are copied onto hard disks and DVDs;
- The DVDs are submitted to the in-charge of scanning centre;
- The hard disks are brought to the central server at the RMSC or DLI project site at IISc for Web enablement.

Structured and admin metadata are entered in online/offline metadata entry form. For this activity, metadata software is used which is produced by DLI national coordination office. This software automatically generates XML file, incorporating Dublin core metadata elements, for each scanned book.

During scanning process, the books are scanned page-wise, one or two pages at a single exposure. The image of each scanned page is captured in a TIFF file, where TIFF stands for Tagged Image File Format – a loss-less uncompressed image file format, widely used in digitization work. Original TIFF (OTIFF) file of each page is preserved for future use. Then image of each page of a book is processed or edited at the scanning centre to comply with quality control policy of DLI. The edited image file is then saved in another TIFF file, called processed TIFF (PTIFF). Scanned images of books written in European scripts such as in English, French, German, and Spanish, are usually converted into editable text formats using optical character recognition (OCR) software ABBYY FineReader. This software delivers superior accuracy (sometimes 99%) in converting any un-distorted image file into texts. Output files from the OCR process are **.**HTML (simple Web page), **.**TXT (simple Text file), and **.**RTF (Rich Text Format).

The process of optical character recognition (OCR) is depicted in Figure 2.27, where the OCR engine corrects image from any distortion; segments texts into paragraphs, lines, words and characters; extract features such as character height, width, horizontal lines, vertical lines, and slope lines; classifies characters based on SVM (support vector machine) algorithm; performs Unicode mapping; and finally recognizes texts in the page. The latest version of OCR software

delivers superior accuracy in converting images into texts, due to continuous innovation in the development of OCR engine.

For each book a unique folder is created to store all files associated with this book. An .XML file is created to describe the book with Dublin core metadata elements. This .XML file also contains admin metadata associated with scanning work of the book. This .XML file is stored in the main folder of the book. Each scanned page is saved as a TIFF image and compressed using the CCITT 4 Fax Compression scheme. Each scanned page is named as an eight digit number such as 00000001.tif and stored in OTIFF folder of the book. The processed image of each page, after necessary quality checking and image editing, is stored in PTIFF folder of the book, having same file name as in OTIFF folder. Similarly, HTML, RTF and TXT folders of the book contain OCR text of each page in the particular file format. The contents of the book are organized in a tree structure as illustrated in Figure 2.28.

The compliance of image quality as per accepted norms for each individual scanned page is an important activity in the quality control process. DLI quality compliance team has identified a number of quality checking parameters that include both manual checking parameters and automatable parameters. Some of the image parameters that can be checked during quality checking process are indicated bellow:

Manual Checking Parameters:

- Skew (twisted image at an angle. De-skewing process is followed where the computer detects and corrects the skew in the image file.)
- Noise (a random, usually unwanted, fluctuation of pixel values in an image)
- Stretch Text (a distorted stretch in text image)
- Flying Text (overlapping or displacement of text image)
- Margins (extra margins in the page)
- Cut text (full image of a page is not captured)
- OTIFF & PTIFF image variation

- Missing pages
- Missing of image files
- Blank Pages

Automatable Parameters:

- Resolution of the Image (conformity to scanned image resolution norm)
- Compression (conformity to CCITT 4 fax compression)
- Size of the TIFF image
- Naming convention (eight digit file name in an arithmetic series)
- Number of files
- Metadata file (whether some metadata elements are missing or incomplete).

Table 2.3 provides a list of major quality parameters for processed TIFF images (Pratha, Ambati & Sankar, 2005). DLI upholds these parameters as a baseline for maintaining uniformity within the project. These parameters are easily achievable utilizing existing of hardware, software and manpower supports available at any DLI scanning centre.

Table 2.3: Major Quality Parameters for processed TIFF images

Name of Parameter	Value
Dimensions	Same size (height x width)
Resolution in dpi (dots per inch)	600 or above
Compression algorithm	CCITT 4 Facsimile Compression
Margin	300 pixels on each side of page
Skew	Less than 2^{o}
Blank pages	Should be identified and annotated
Source: (Pratha, Ambati & Sankar, 2005)	

Figure 2.26 identifies quality assurance processes at different stages of scanning activity that include internal quality control as well as external quality control mechanisms. At each and every stage of scanning work, there is a check in quality. Different parameters for image quality compliance are checked in some

stages of quality control. Then final digital objects are produced within an acceptable quality.

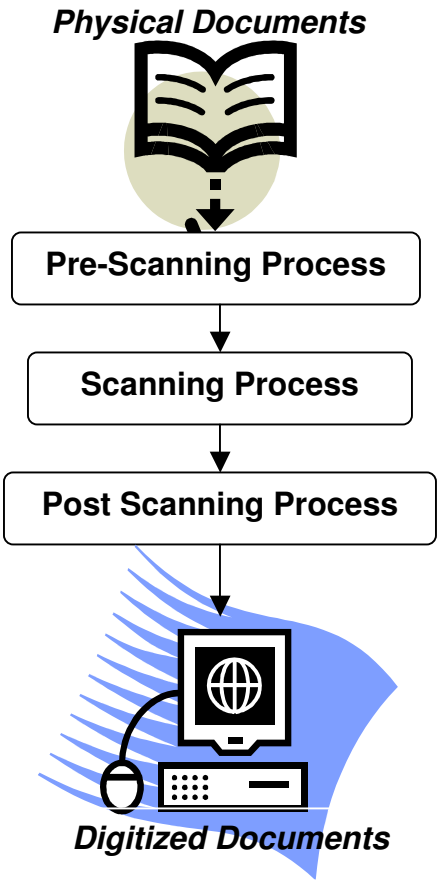


Fig. 2.22: DLI Process Workflow

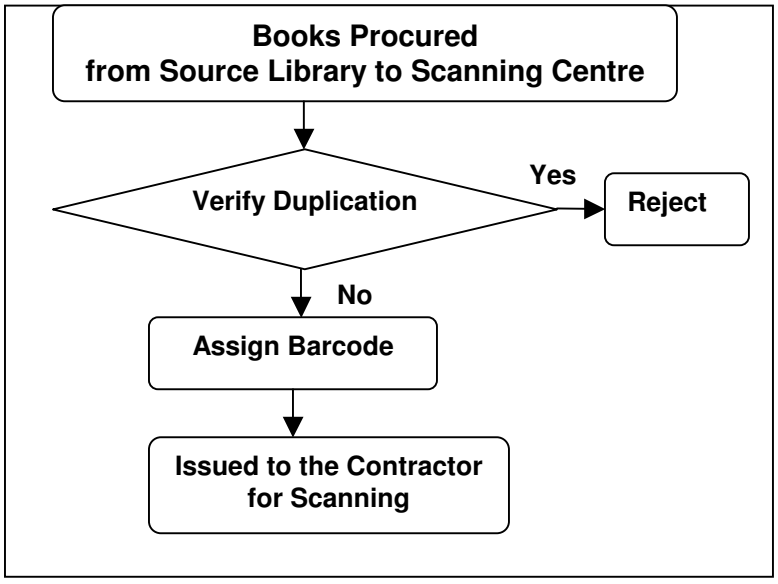


Fig. 2.23: Workflow in Pre-scanning Process

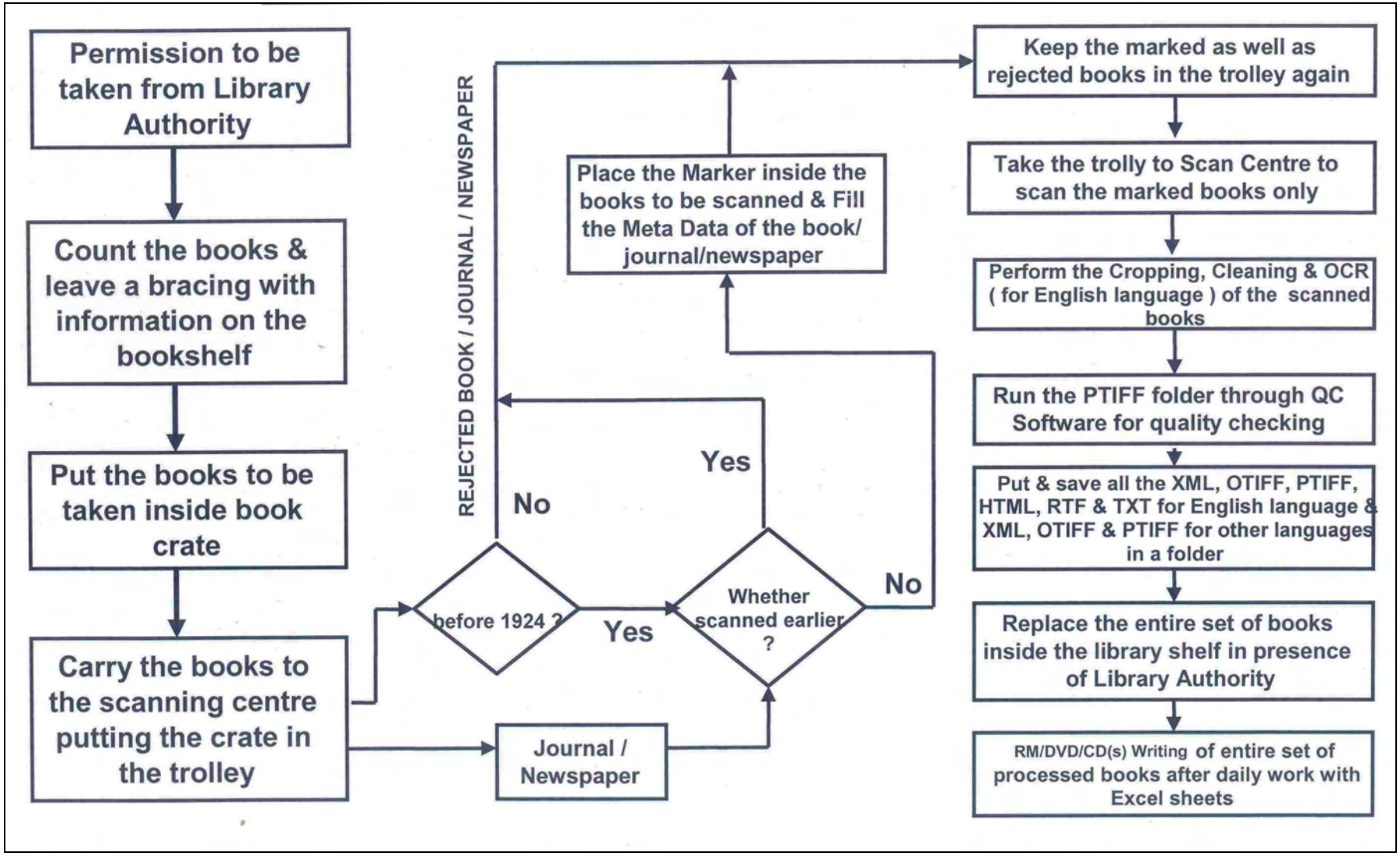


Fig. 2.24: Book Movement within the Source Library and Scanning Centre

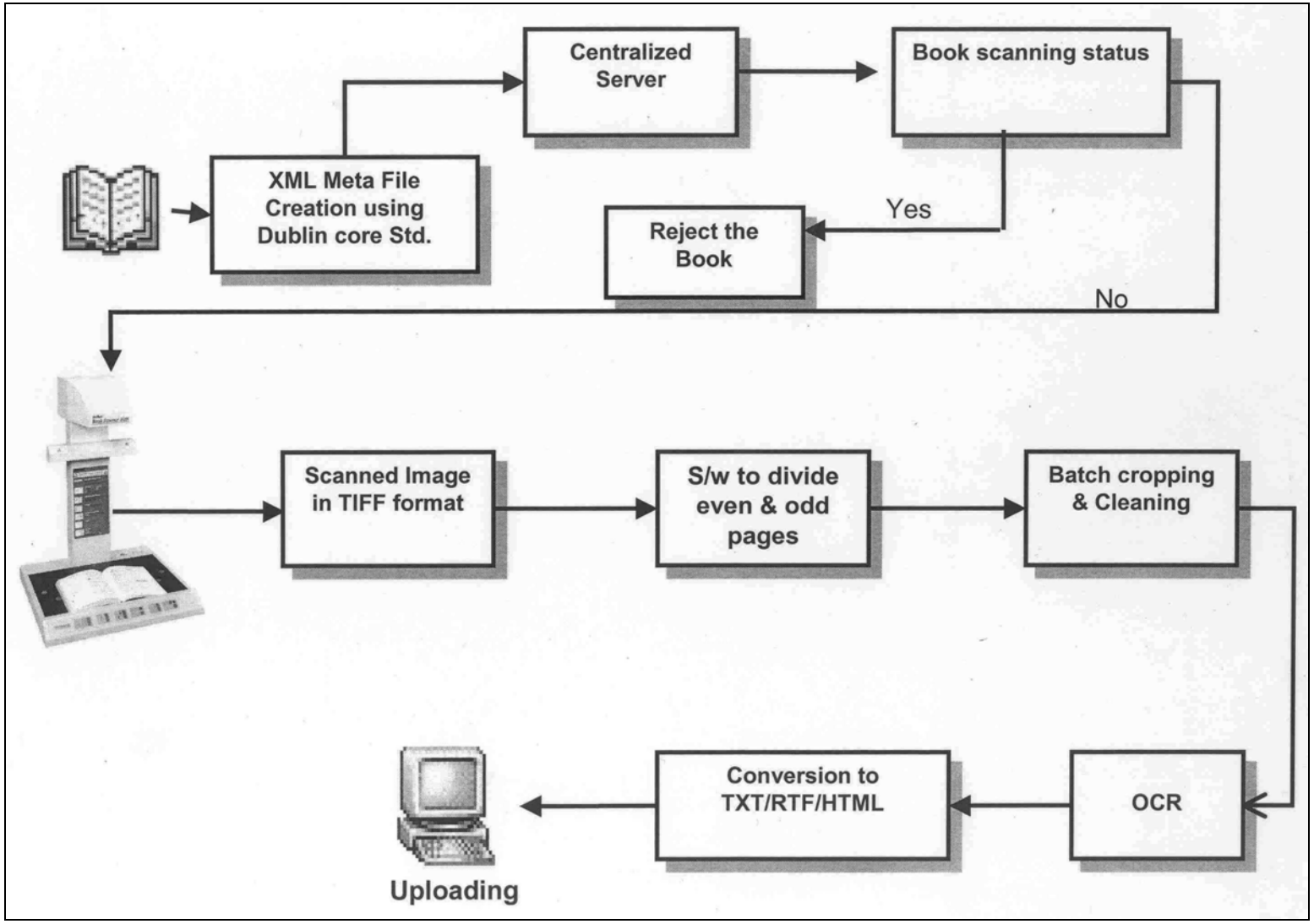


Fig. 2.25: Process of Digital Preservation in Scanning Centre

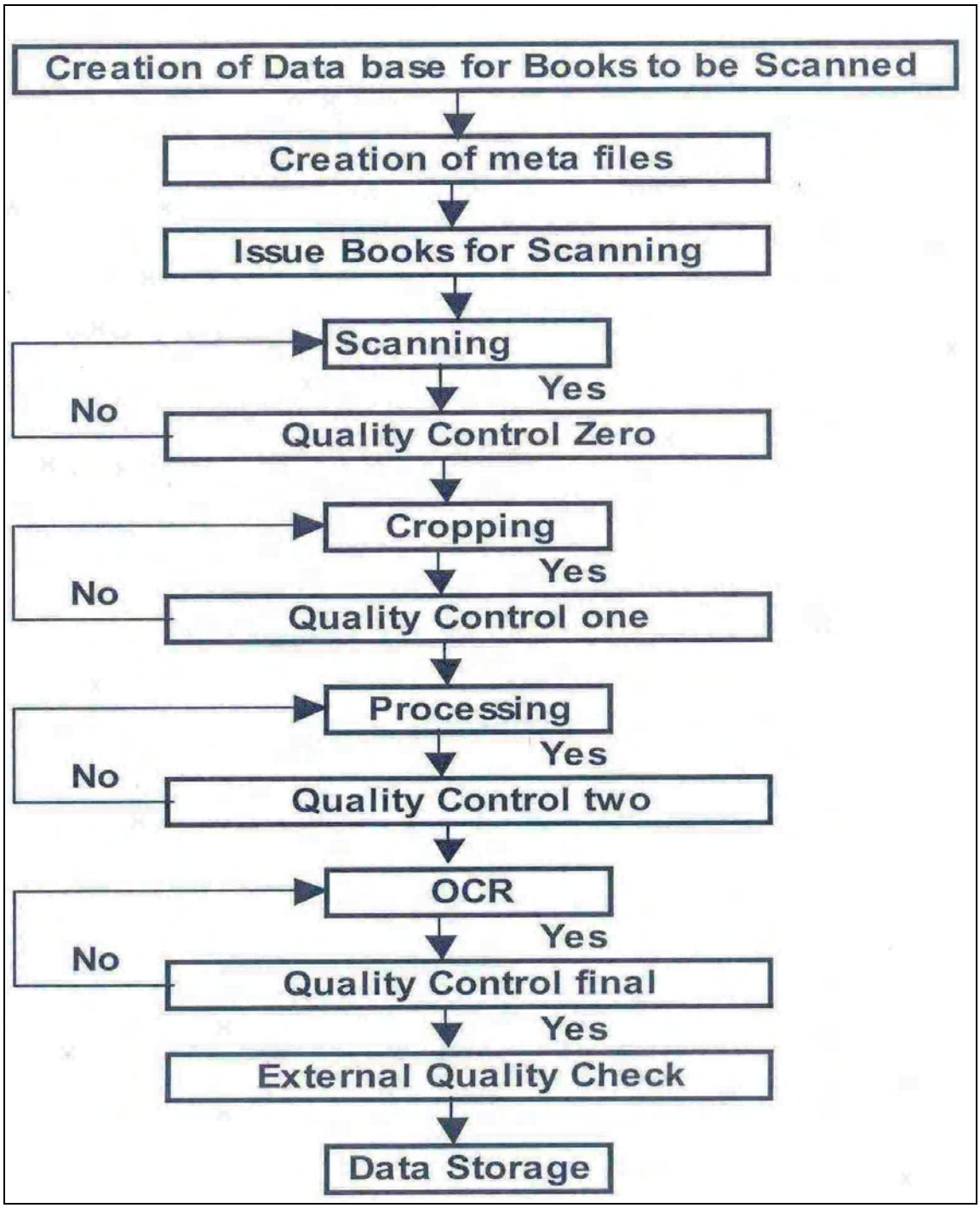


Fig. 2.26: Quality Control Processes at Scanning Centre

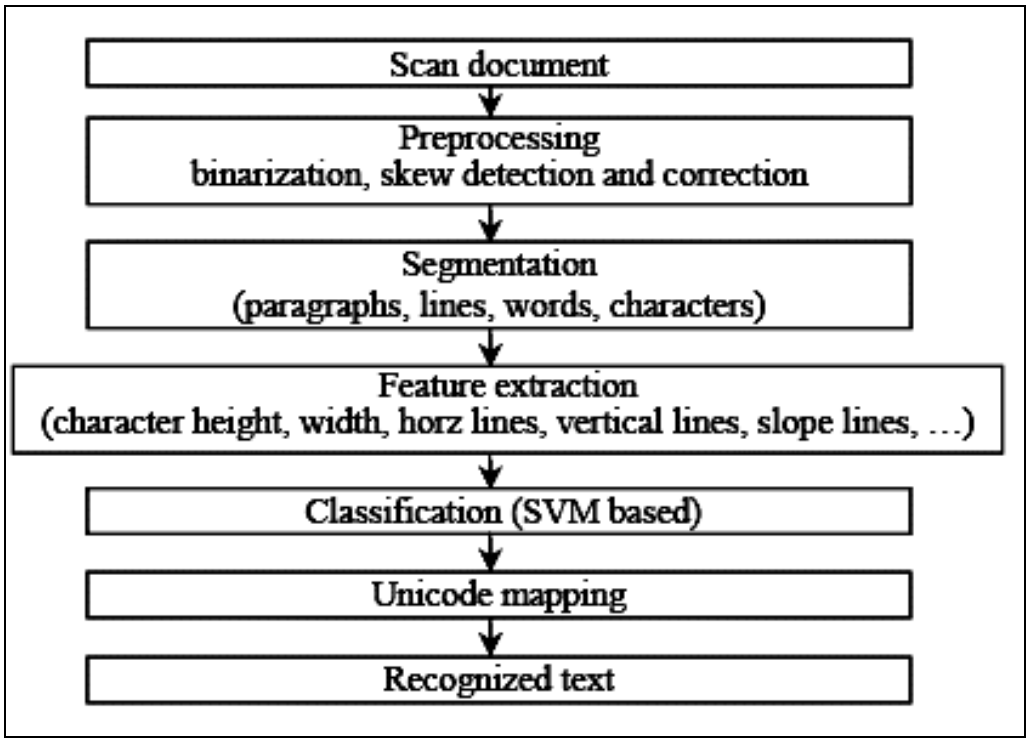


Fig. 2.27: Optical Character Recognition (OCR) Process

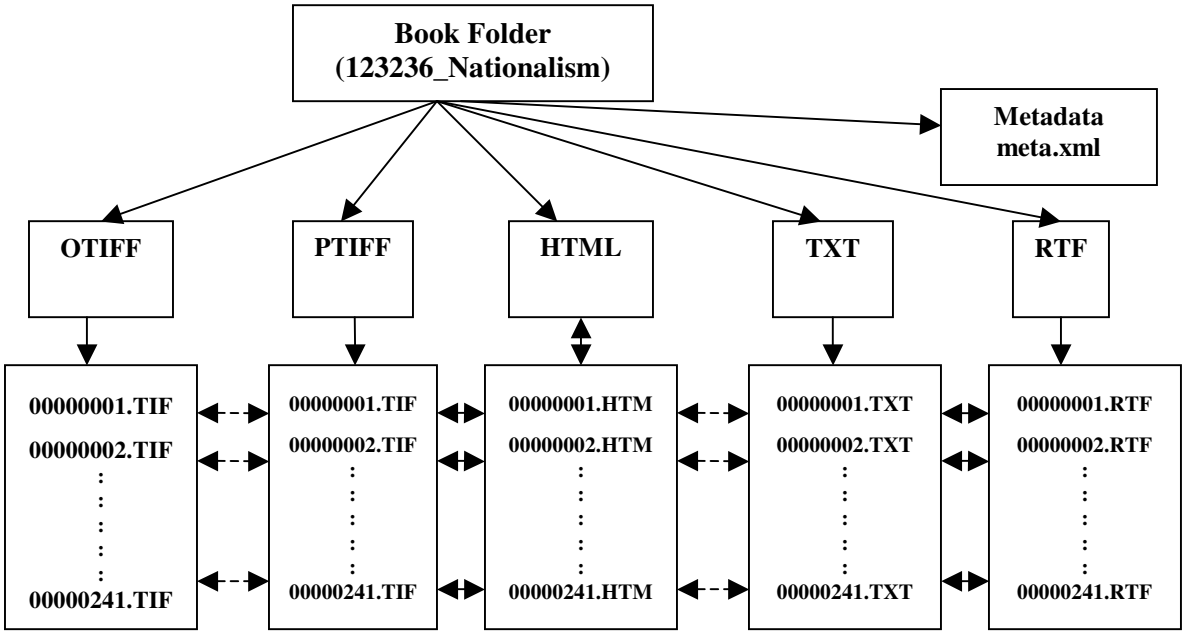


Fig. 2.28: Tree Structure of Contents of a Digitized Book

2.9 Scanning Equipments Procured and Used in DLI Project

A scanning unit comprises a book scanner, a workstation equipped with some essential software. A book scanner is essential equipment in a DLI project that completes a scanning setup. In DLI project two types of book scanners are usually used, namely, Konica Minolta Book Scanner (series PS6500 and PS7000) and OmniScan Zeutschel Book scanner (series 7000). Both kinds of book scanners use a book cradle and overhead scanning system. These scanners facilitate face-up scanning – the ability to scan valuable books without too much handling. There is no need to turn volumes over in face-down position, as in the case of desktop scanners. These scanners can automatically compensate for page curvature and can eliminate centerline shadows, outside frames, even fingers holding the outside edges of pages. For books with fragile bindings, scanner's book table bends lower in the centre or on either side to accommodate the spine. These book scanners also have large format scanning area – scanning area is big enough to accommodate oversized bound volumes, manuscripts, archival records, artworks, cartographic and legal documents with ease. These are meant to be easy for anyone to use and seem to offer features to help those who are less skilled. These book scanners are fast, safe, and simple to use. The book scanners also provide high image quality - clean images with 600 dpi or more resolution.

Figure 2.29 provides a sketch of manual book scanner, specifically OmniScan Zeutschel book scanner that captures images using a digital camera and wide-angled light sources. Figures 2.30 and 2.31 provide images of different brands of book scanners used in DLI project.

Figure 2.30 also shows a workstation, where book scanner is connected. In this workstation scanning software such as ABBYY FineReader or OmniScan is installed. The book scanner produces high quality images, and scanning software processes those images and stores in hard disk of that workstation.

Procurement of these book scanners leveraged overall productivity and efficiency of scanning centres across India. These scanners also helped project leaders to realize some of the project goals.

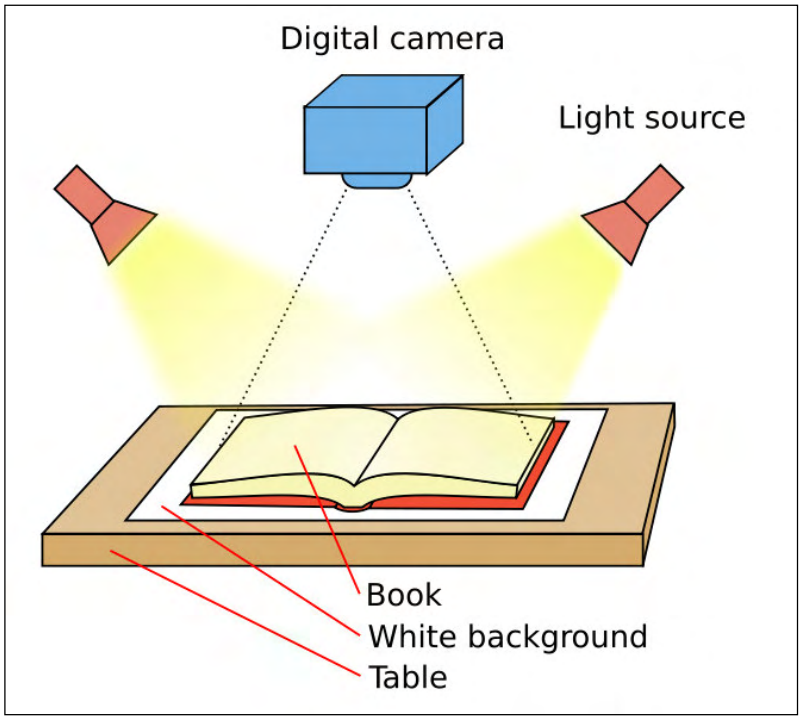


Fig. 2.29: Sketch of a Manual Book Scanner

Fig. 2.30: Konica Minolta Book Scanner Equipped with Workstation

Fig. 2.31: OmniScan Zeutschel Book Scanner

2.10 Software Developed, Customized and Used in DLI Project

Indian Institute of Science (IISc) Bangalore and International Institute of Information Technology (IIIT) Hyderabad are research partners of UDL project. They have collaboratively developed several software tools, which are now used in different stages of digitization workflow, particularly in regional mega scanning centres and scanning centres. Some of the software tools were developed by UDL project partners abroad. Later those tools were customized to suit the requirements of Indian partners. IISc Bangalore, in association with IIIT Hyderabad, took leadership in the customization activity for the benefits of Indian partners. On the other hand, some software tools are adopted and used by UDL project partners abroad.

During the scanning process, structure metadata and admin metadata are entered in scanning centre to provide a detailed profile of scanned book. A metadata software is thus developed to cater to the purposes of providing detailed information on book such as author, title, publisher, year of publishing, etc. and scanning work such as name of scanning centre, barcode assigned,

name of digital re-publisher, etc. This metadata software provides an interface of metadata entry form.

Many Indian books being scanned in scanning centres across India are actually written in Indian languages. In metadata entry form there should be a facility to enter metadata in Indian languages, while books are actually written in Indian languages. For example, title and author of a Bengali book should be entered in that language and should have provision of retrieving same data in that language. Metadata software used in DLI project supports Indian fonts based on UNICODE for Indian languages. In metadata software there is a provision to enter metadata in an Indian language and in English. For systematic representation of name of authors or book title in an Indian language into English text, transliteration is needed. DLI has produced (i) Om Transliteration Mapping Scheme for Indian Languages, and (ii) Om Transliteration Editor, to transliterate Indian texts into English and vis-à-vis. Om Transliteration Editor is an Indian language text editor based on Om Transliteration Mapping scheme. This is free software and can be downloaded from <http://swati.dli.ernet.in/om/> for non-commercial use. This software supports Indian fonts for few major Indian languages. This software maintains a mapping table for Indian languages that helps in converting Indian alphabets into English texts. DLI has published an iTrans manual that is very much useful to the new users and metadata-entry operators in DLI scanning centres. Figure 2.33 provides some examples of transliteration from Devanagari script, using Om Transliteration Editor.

In earlier section, it was described that in the pre-scanning process duplication of book is checked to avoid the same book repeatedly digitized by other scanning centres. A duplication checking software tool is thus developed for this purpose. Similarly, a comprehensive Quality Assurance Software Package is developed and used across the RMSCs and scanning centres, that includes (i) Duplication Checking Tool, (ii) Image Quality Checking Tool, and (iii) Metadata Quality Checking Tool.

DLI project has developed several other utility software tools for the convenience of system administrators in the project sites. These software tools are: (i) Server Management Tool, which provides server-site applications and maintains its services; and (ii) Workflow Management Software, which is monitoring and evaluation tool to determine any bottleneck or any deviation from targets and deadlines.

A glimpse of software tools developed, customized and used is shown in Figure 2.32. *OM i-Trans Editor*, a transliteration tool for Indian languages, is an example of many useful software tools developed in this project. This software is shown in Figure 2.33.

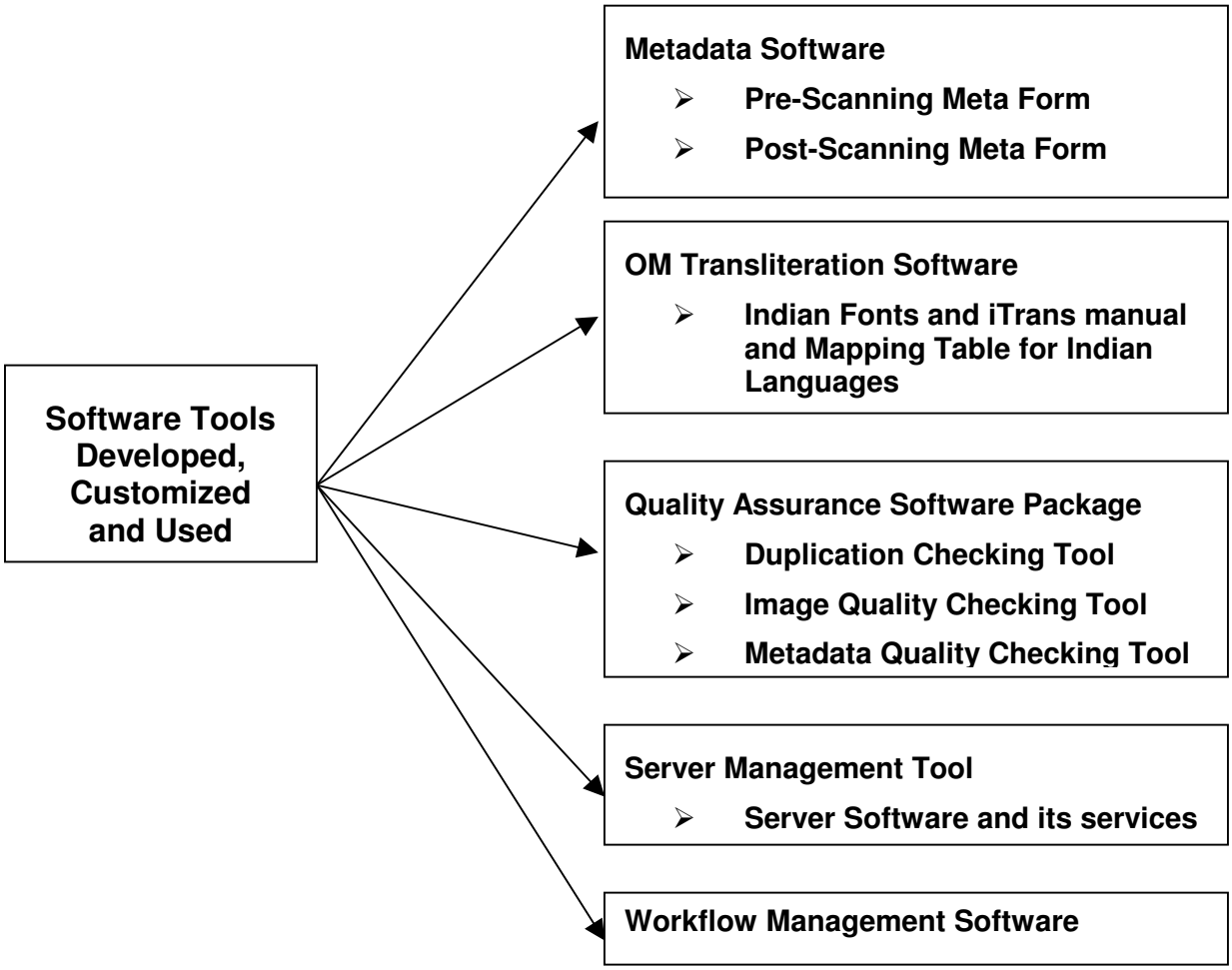


Fig. 2.32: Software Tools Developed and Used in DLI Project

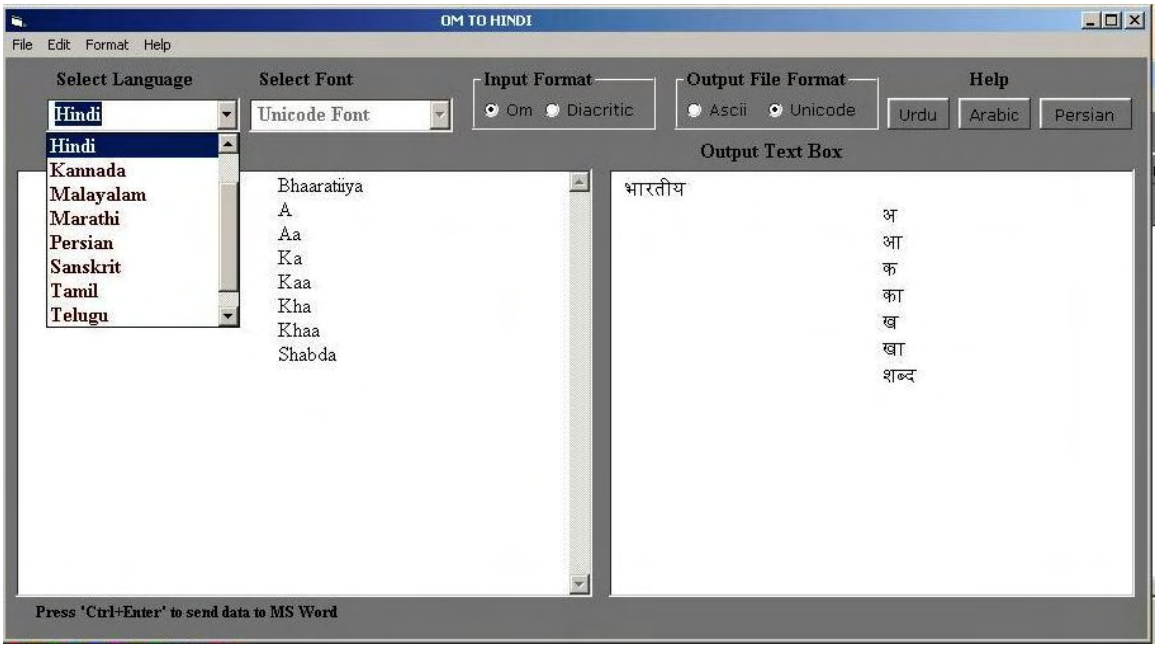


Fig. 2.33: Indian Language Transliteration using OM i-Trans Editor

2.11 Metadata Used in DLI System

Metadata is the data for describing physical or digital documents in ways that make them easier to find. A metadata schema provides a simple and standardized set of conventions for describing resources online. Metadata is also being considered as the brain of any online retrieval system, including a digital library system. Many metadata schemas are in existence, where different sets of elements are used for describing different kinds of electronic resources. Dublin core metadata element set is the most popular standard for cross-domain information description in a digital library system.

The Extensible Markup Language (XML) facilitates the sharing of structured data across different information systems, particularly via the Internet. Structured metadata describing electronic resources are embedded in XML documents to make the digital library system interoperable and to allow crosswalk.

Digital Library of India initiative is digitizing several thousands of physical documents in several scanning centres across India. DLI Web-enabled those digitized documents to make them available online through the DLI portals. The

searching or navigating of electronic resources from a large pool became possible due to proper representation of those documents in different metadata elements. For each scanned book, DLI maintains regular metadata using Dublin Core metadata element sets that embedded in an XML file. Structured and Administrative metadata are also provided for each book, where many elements describe the book and its scanning work.

Table 2.4 identifies metadata elements usually entered into DLI system using an online/ offline metadata entry form, where both structured and admin metadata elements are entered by the metadata-entry operators in the scanning centres. This Table also shows which metadata elements are mandatory. Online metadata entry form also facilitates some guidelines to un-experienced metadata-entry operators on how to render metadata elements in the system. For the convenience of metadata-entry operators, some metadata elements can only be entered by selecting from a given list. This selection process also reduces chance of wrong metadata entry.

Figure 2.34 illustrates metadata structure and metadata elements used in describing a scanned book. This Figure also identifies the metadata elements that are part of regular metadata, structural metadata and admin metadata. These three sets are strategically displayed separately in DLI portals to make them easy to understand/ grasp.

Figure 2.35 provides a glimpse of structural metadata of a retrieved book that provides bare minimum elements that describe a book titled *Nationalism*, written by Rabindranath Tagore. Figure 2.36 provides a glimpse of administrative metadata of the same book along with some details on the scanning work, e.g., name of scanning centre, name of source library, date of scanning, digital re-publisher, etc.

The metadata elements, available in administrative metadata as well as in structural metadata, are also tagged in XML file for the book. On the other hand it can be said that XML sheet is one of the source documents available with DLI storage system to generate structured and admin metadata which displayed on

screen. Figure 2.37 shows tagged metadata displayed in XML file for the book titled *Nationalism*, written by Rabindranath Tagore.

In the scanning centre, some metadata fields are not filled up due to many reasons. These unspecified metadata elements are also displayed blank on screen. But, key metadata elements are always filled up and checked during different phases of quality control.

In Digital Library of India portals, some metadata elements are searchable whereas some others are not searchable. Searchable metadata elements are:

- Title of the Book
- Name of Author
- Subject
- Year of Publishing
- Language, and
- Name of Scanning Centre.

Search result produces a list of books matching the search criteria, where following metadata are displayed:

- Title of Book
- Name of Author
- Subject
- Language
- Year of Publishing, and
- Total number of pages.

The Advanced Search Interface is available in the DLI homepage (Figure 2.38). This interface also specifies the searchable metadata elements. In advanced search, combination of two or more elements to narrow the search is possible.

Fig. 2.34: Metadata Used in Digital Library of India

Table 2.4: Metadata Elements in Online Metadata Entry Form

Name of Metadata	Whether Mandatory Field	Example & Help
Language	Yes	Choose one from list: Assamese; Bengali; English; Gujarati; Hindi; Kannada; Malayalam; Marathi; Oriya; Persian; Punjabi; Sanskrit; Urdu; Tamil; Telugu; Other. If Other, Please enter the input in the text box
Title	Yes	Should be entered in Title Case. E.g., A System of Physical Chemistry; Applied Nuclear Physics. ITrans Input for UNICODE text allowed.
Creator/Author	Yes	E.g., Raman, C.V. Sir (for Sir C.V.Raman); Smith, John ed (for John Simth ed). ITrans Input for UNICODE text allowed.
Keywords Description	No	
Subject	Yes	Choose one from list: Generalities; Philosophy/ Psychology; Religion/ Theology; Social Sciences; Natural Sciences; Technology; the Arts; Language/ Linguistics/ Literature; Geography/ Biography/History; Vacant. May choose a sub-category.
Publisher	Yes	Enter the Publishers Name as it appears. E.g., John Wiley; Cambridge University Press.
Contributor	No	Enter, if there is an Editor, Illustrator, Translator etc., in addition to creator.
Date	Yes	YYYY/MM/DD format E.g., only Year is known, e.g. 1923/00/00 Both Year and Month are known, e.g. 1924/03/00 All three entries are known, e.g. 1926/11/26
Document Type	Yes	Choose one from list: Art Objects; CDs, DVDs, Film Reels; Fabric; Floppies;

		Glass; Magnetic Tape; Microfiche; Microfilm; Palm leaf; Paper; Wood. Category of the resource i) Physical type ii) Machine Readable Form iii) 3-D images.
Format	No	TIFF
Identifier	No	URL http://www.nameofsite.org/ ISBN 0304701564 Accession No. DLI334455 (when DLIB India Number is assigned)
Source	Yes	Carnegie Mellon University [the resource is from Carnegie Mellon University]
Relation	No	Relationship of this resource. E.g., [Is a textual version of an audio]; [Is Part of: Is contained in another resource]
Coverage	No	
Rights	No	Choose from list: Copyright permitted; In Public Domain; Not Available
Copyright Date	No	Same as Publication Date unless copyright date is different from the Publication Date or write beyond copyright restriction.
Scanning Centre	Yes	Choose one from list.
Scanner Number	No	
Digital Republisher	Yes	Digital Library of India (Name of the Centre Scanning the document)
Digital Publication Date	Yes	Choose one from list. Date on which the document is Scanned.

Title	Nationalism
Author1	Tagore, Rabindranath
Author2	
Subject	PHILOSOPHY. PSYCHOLOGY
Language	english
Barcode	4990010103875
Year	1918
BookReader-1	Click here
BookReader-2	Click here

Admin Metadata	**Structural Metadata**

Fig. 2.35: Structural Metadata displayed for a Retrieved Book

Title	Nationalism
Author1	Tagore, Rabindranath
Author2	
Subject	PHILOSOPHY. PSYCHOLOGY
Language	english
Barcode	4990010103875
Year	1918
Format	Tagged Image File Format
Publisher	Macmillan and Co., London
Vendor	NONE
Scanning Centre	cdack
Scanning Location	NONE
Source Library	RAMKRISHNA MISSION INSTITUTE OF CULTURE, GOLPARK
Digital Re-Publisher	Digital Library Of India
Digital PublicationDate	2006-08-22
Numbered Pages	
Un-Numbered Pages	
TotalPages	152
Table of Contents	Click here
Read Online	Click here

Fig. 2.36: Administrative Metadata displayed for the Same Book

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <dublincore>
  <language>English</language>
  <title>Nationalism</title>
  <title1 />
  <title2 />
  <title3 />
  <creator>Tagore, Rabindranath</creator>
  <creator1 />
  <creator2 />
  <creator3 />
  <creator4 />
  <creator5 />
  <creator6 />
  <creator7 />
  <creator8 />
  <keywords>Nationalism - India</keywords>
  <keywords1>Nationalism - Japan</keywords1>
  <keywords2 />
  <keywords3 />
  <keywords4 />
  <description />
  <subject>PHILOSOPHY. PSYCHOLOGY</subject>
  <subject1>Specific Philosophical Systems and Viewpoints</subject1>
  <subject2>Nationalism</subject2>
  <publisher>Macmillan and Co., London</publisher>
  <contributor />
  <contributor1 />
  <date>1918/00/00</date>
  <documenttype>Print - Paper</documenttype>
  <documenttype>Book</documenttype>
  <format>Tagged Image File Format</format>
  <identifier />
  <source>RAMAKRISHNA MISSION INSTITUTE OF CULTURE,
GOLPARK</source>
  <relation />
  <coverage />
  <rights>In Public Domain</rights>
  <copyrightdate />
  <scanningcentre>cdack</scanningcentre>
  <barcode>04990010103875</barcode>
  <scanningnumber>RMIC_02</scanningnumber>
  <digitalrepublisher>Digital Library Of India</digitalrepublisher>
  <digitalpublicationdate>2006/08/22</digitalpublicationdate>
  </dublincore>
```

Fig. 2.37: Tagged Metadata Displayed in XML Format for the Same Book

Fig. 2.38: Advanced Search Interface, available in DLI Homepage

2.12 Information Retrieval in Digital Library of India Portals

Digital Library of India portals provide convenient online information search and retrieval facility to users across the world. DLI portals also provide easy navigation interface, where users can conveniently navigate through many browsing parameters.

Browsing the Collections

Users can browse the collections to view the books in alphabetically orders. Collections can be navigated by Book's Title, Author's Surname, Subject, Language and range of Years.

By Title: User can click on any alphabet from A to Z to view all the book titles starting from that alphabet.

By Author: User can click on any alphabet from A to Z to view all the Author surnames starting from that alphabet.

By Subject: User can select a subject from given list of the subjects (e.g., Architecture, Art, Astronomy, Autobiography, Commerce, Devotional, Economics,

Electrical, Engineering, General, Geography, Health, Law, Literature, Mathematics, Medical, Mythology, Philosophy, Poetry, Politics, Religion, Sciences, Sociology, Statistics and Zoology).

By Language: User can select a language from given list of the languages (e.g., Bengali, English, Hindi, Kannada, Marathi, Sanskrit, Tamil, Telugu and Urdu).

By Year: User can select the a range of years from given in the list of ranges (e.g., 1850-1900, 1901-1910, 1911-1920, 1921-1930, 1931-1940, 1941-1950, and 1951 onwards).

Figure 2.39 illustrates collect navigation facility given in the DLI homepage. This Figure also indicates that collection navigation for few scanning centres is also possible from this interface.

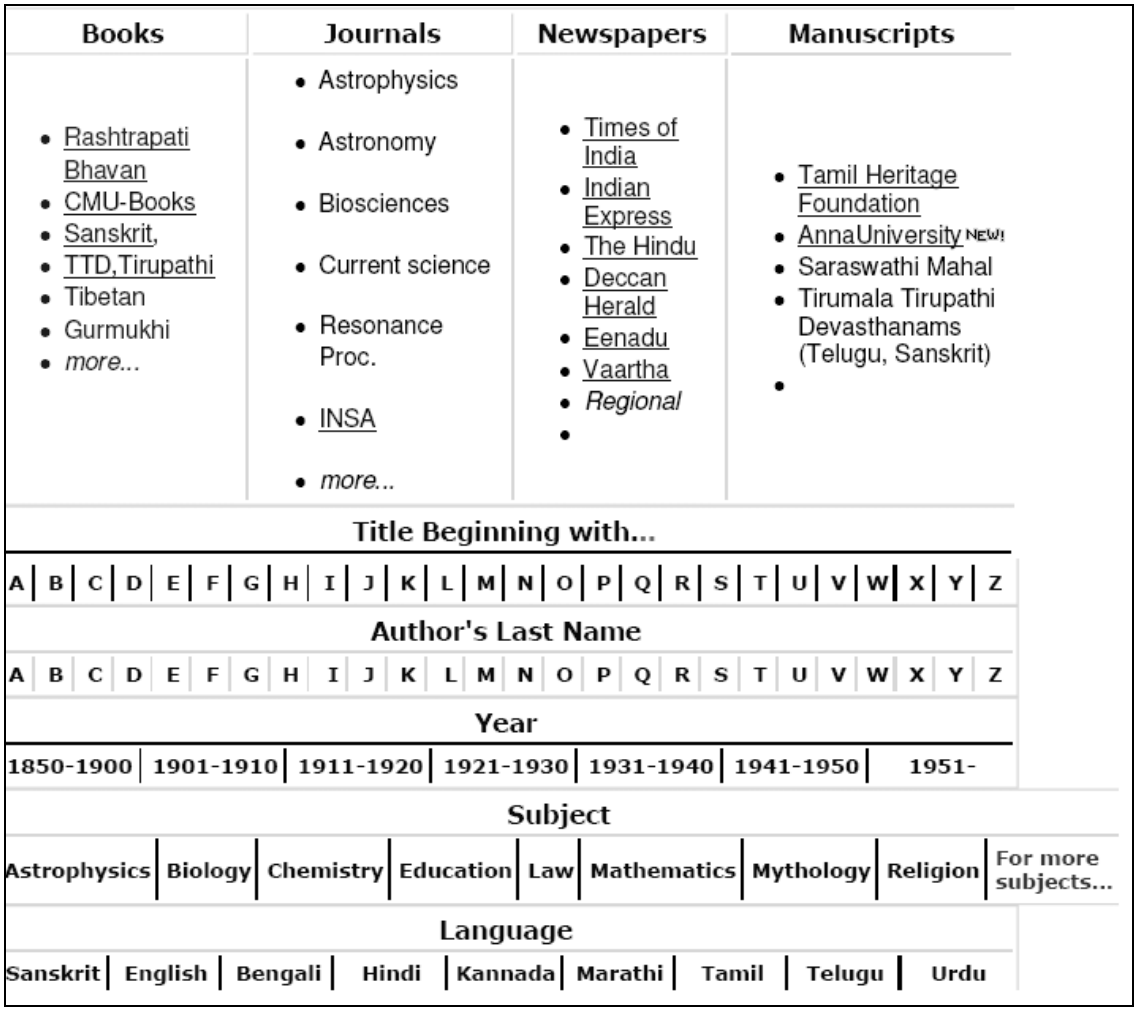


Fig. 2.39: Browsing of Collections from Homepage of Main DLI Portal

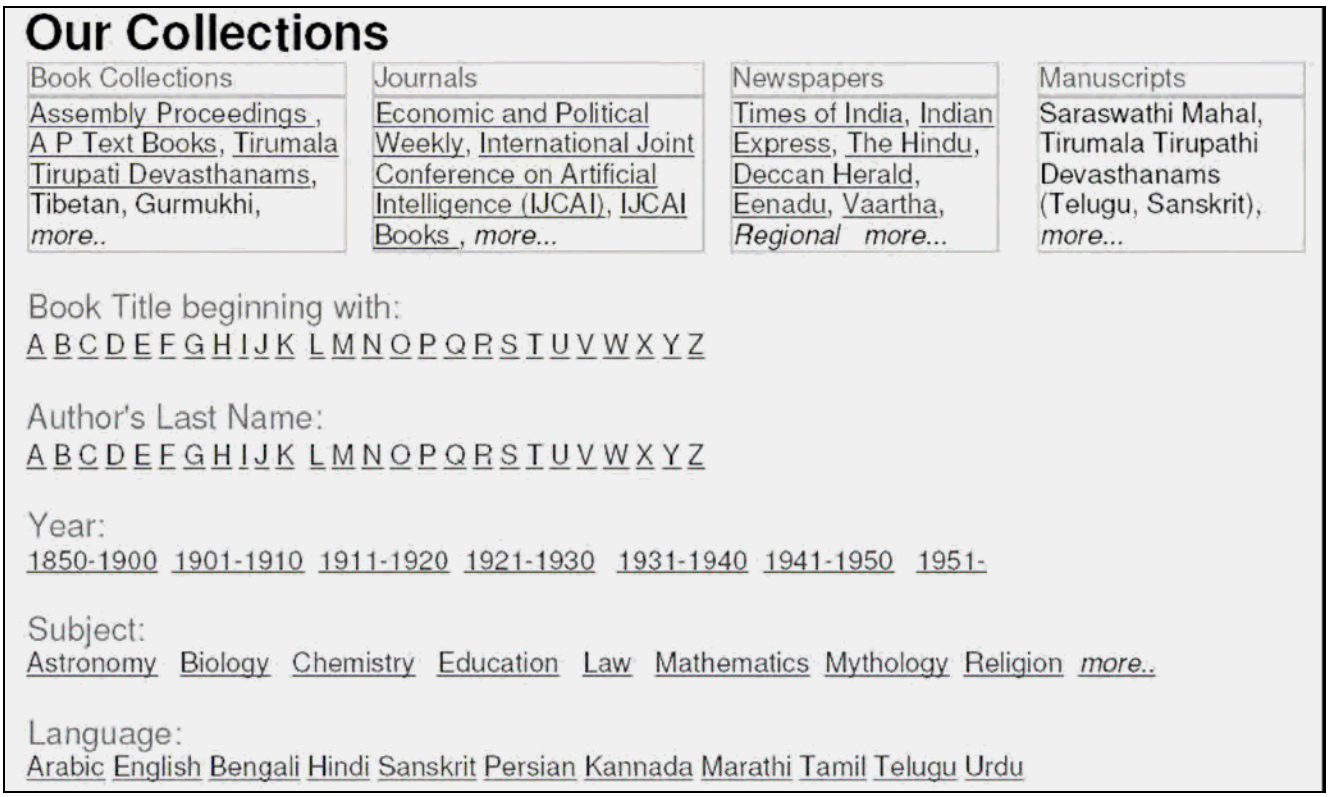


Fig. 2.40: Browsing of Collections from Homepage of DLI Portal of RMSC Hyderabad

Searching the Collections

By Title: A user can search book titles from DLI homepage. For this purpose, user can enter a word, a phrase, or a subset of the word of title to search the book. The search returns all books in the DLI system whose title matches the words user entered in the search box. Each search word must have at least 4 characters length. User can enter words or subset of words in any supported language (English, Hindi, Urdu, etc.), if Unicode-based multilingual support is enabled for the user's computer and if he knows typing the language mentioned using the qwerty keyboard. Otherwise he may use the transliteration tool available offline using OM i-Trans Editor as depicted in Figure 2.33 (offline), available online as depicted in Figure 2.46 (online). These transliteration tools are available for texts in Indian languages and provide the equivalent texts in Unicode.

By Author: User can enter full name, surname, first name or partial name of the author of the book for searching a book. For this field Unicode search is enabled, similar to Title search.

By Subject: User can select a subject from given list of the subjects (e.g., Architecture, Art, Astronomy, Astrophysics, Autobiography, Biology, Chemistry, Commerce, Dance, Devotional, Economics, Education, Electrical, Engineering, General, Geography, Health, History, Law, Literature, Marketing, Mathematics, Medical, Music, Mythology, Other Subjects, Philosophy, Poetry, Politics, Psychology, Religion, Science, Sociology, Statistics, Zoology and Unknown). It is also noticed that in the search interface more number of subjects are listed than browse interface.

By Language: User can select a language from given list of the languages (e.g., Arabic, Bengali, English, French, German, Greek, Gujarati, Hindi, Irish, Italian, Kannada, Marathi, Norwegian, Persian, Russian, Sanskrit, Spanish, Tamil, Telugu, Urdu, Multilingual, and Others). It is also noticed that in the search interface more number of languages are listed than browse interface.

By Year: Here user can type in the year or range of years in this field to search for a book ranging the particular years, e.g., 1909 to 1947. This range of years is user-driven, not the system driven as in browse interface.

By Scanning Center: Here user can select a scanning centre from given list of the scanning centres (e.g., CDAC Kolkata, Tirumala Tirupati Devasthanams, AKCE, SASTRA and many more).

Advanced Search

In the DLI Homepage advanced search interface is available where user can search by more than one search parameters at a time. Users utilize advanced search option to refine a search, i.e., to avoid a large set of data being retrieved using only a single parameter. When users enters two search parameters in the advanced search interface, DLI system automatically takes an AND Boolean

Operand for each added parameter. For example, when 'physics' is given in the subject field, information on all books pertaining to that particular subject is appeared in search result. For further narrowing the search result, another parameter can be given in the same search interface such as Bengali language. Here DLI system interprets it as "Subject=Physics AND Language=Bengali", i.e. physics books in Bengali language to be appeared in the search result. Similarly, few other parameters can also be added for further narrow down a search result, e.g. a year (1950), or a range of years (1900-1950), or a scanning centre (CDAC Kolkata).

Figure 2.41 illustrates how two search parameters are being entered in the advanced search interface. Figure 2.42 shows a shaded area where Boolean operand AND finds books matching both parameters.

Similarly, Figure 2.43 illustrates how three search parameters are being entered in the same advanced search interface. Figure 2.44 shows a shaded area where Boolean operand AND finds books matching all parameters.

Figure 2.45 illustrates advanced search interface available with DLI Portal of RMSC Hyderabad. This Figure also indicates that more searchable metadata elements are available with this portal. DLI main portal provides six searchable metadata fields, i.e., Title, Author, Subject, Year, Language and Scanning Centre. Additional searchable fields available with DLI portal in IIIT Hyderabad are: Vendor, Source Library and Scanning Location.

In advanced search interface, Unicode-based Indian language texts can be entered only in author and title metadata fields. Other metadata fields are enabled in English only, thus, do not accept Indian language search strings. If an end user does not have Indian language typing facility or skill, he can always use an Indian language transliteration tool for getting Unicode-based Indian language text inputs, as illustrated in Figures 2.33 and 2.46.

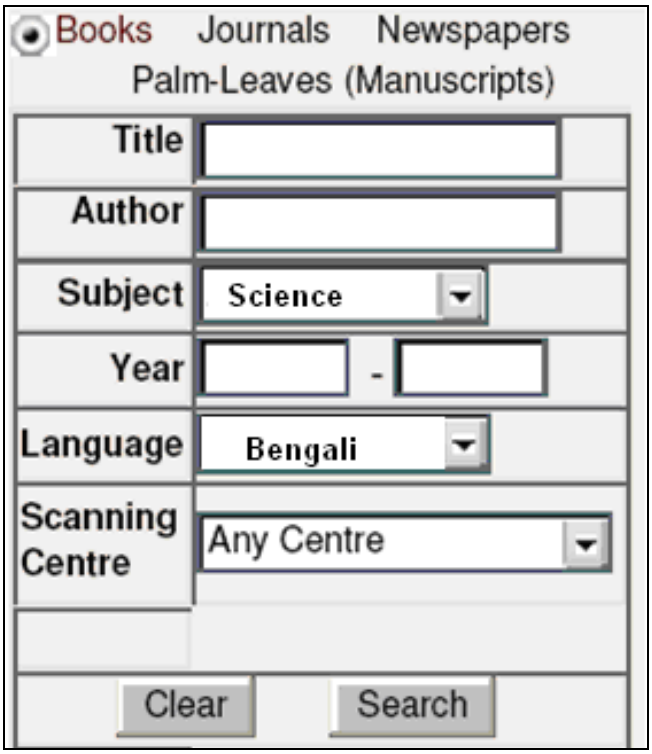


Fig. 2.41: Two Search Parameters Entered in Advanced Search Interface

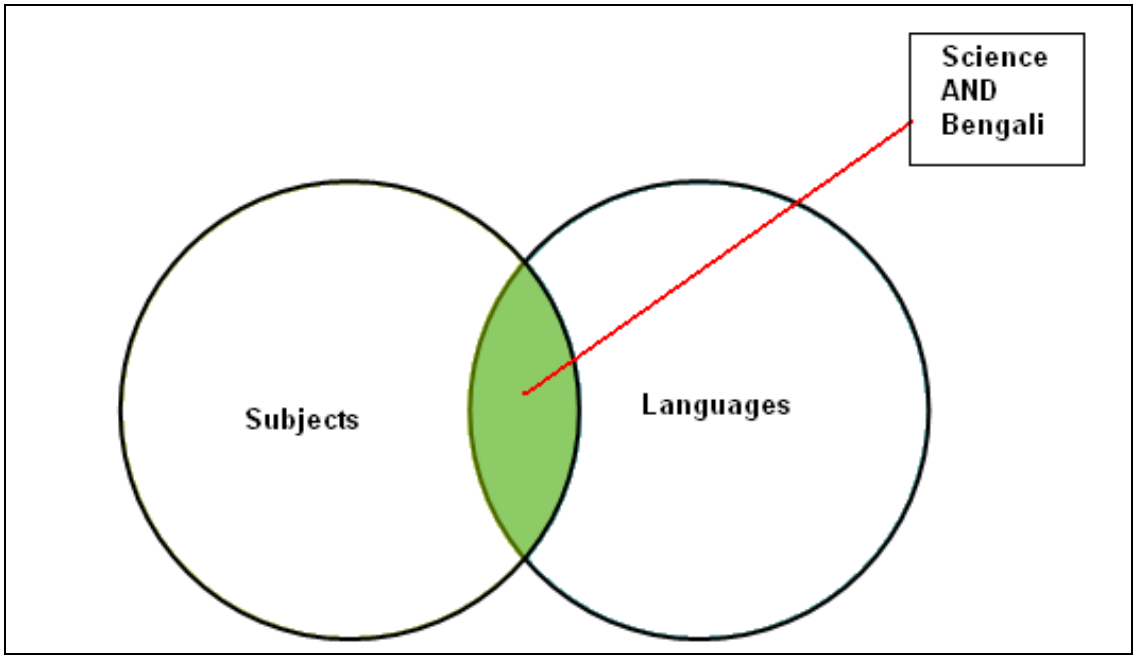


Fig. 2.42: Boolean AND Operand Matching Both Search Parameters

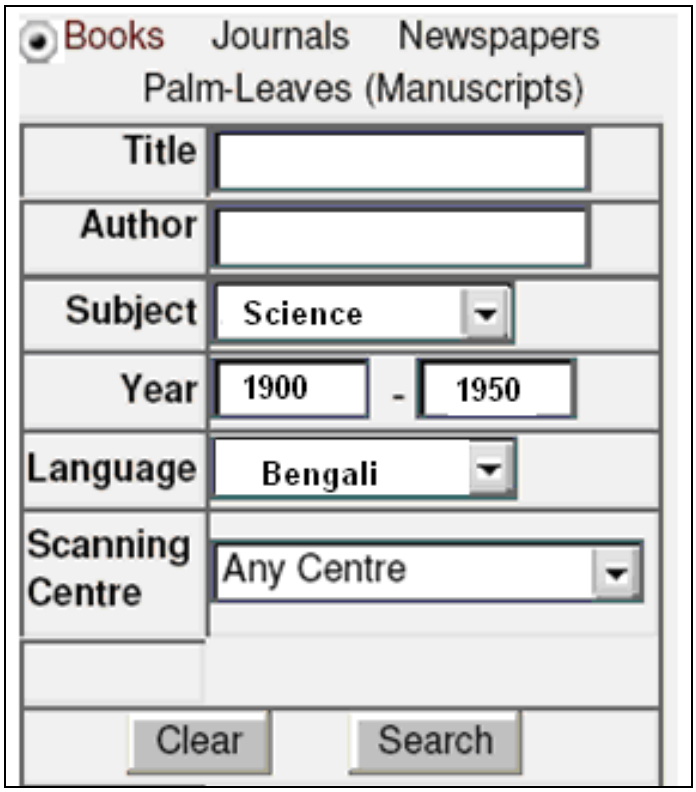


Fig. 2.43: Three Search Parameters Entered in Advanced Search Interface

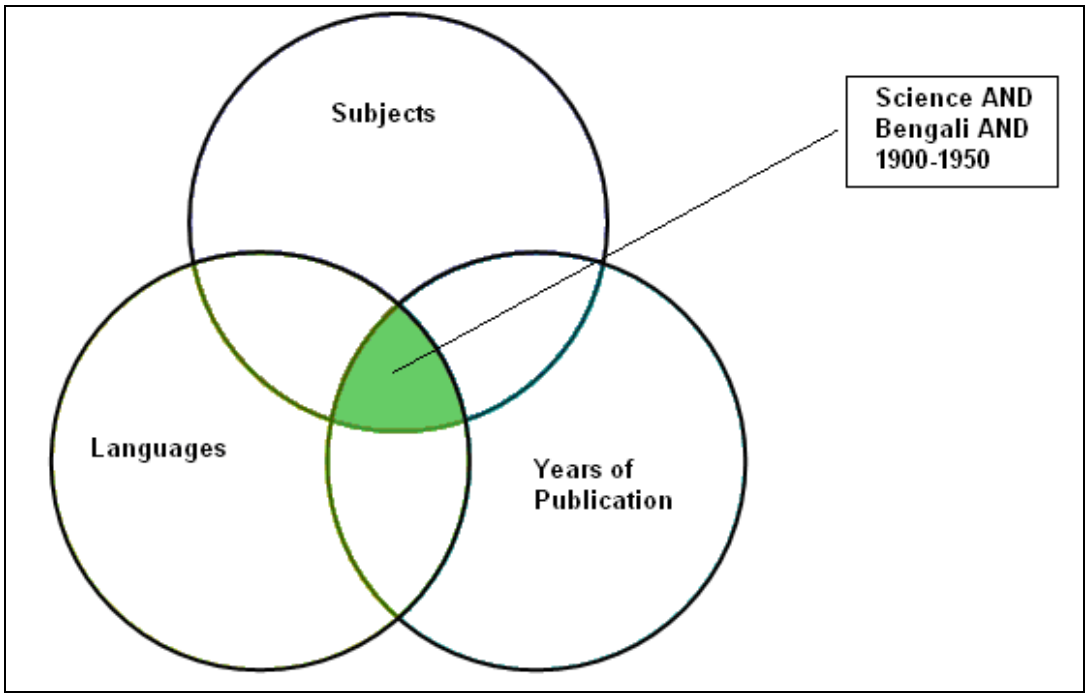


Fig. 2.44: Boolean AND Operand Matching All Search Parameters

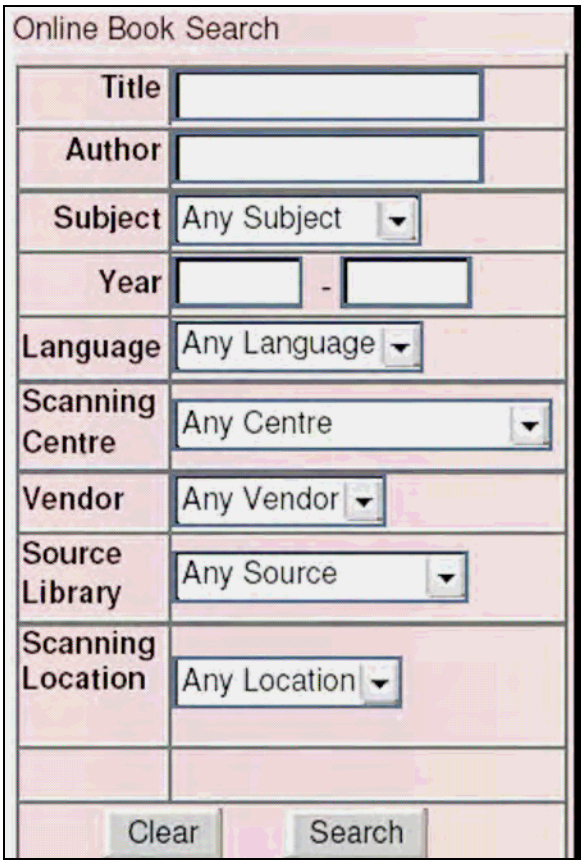


Fig. 2.45: Advanced Search Interface at DLI Portal of RMSC Hyderabad

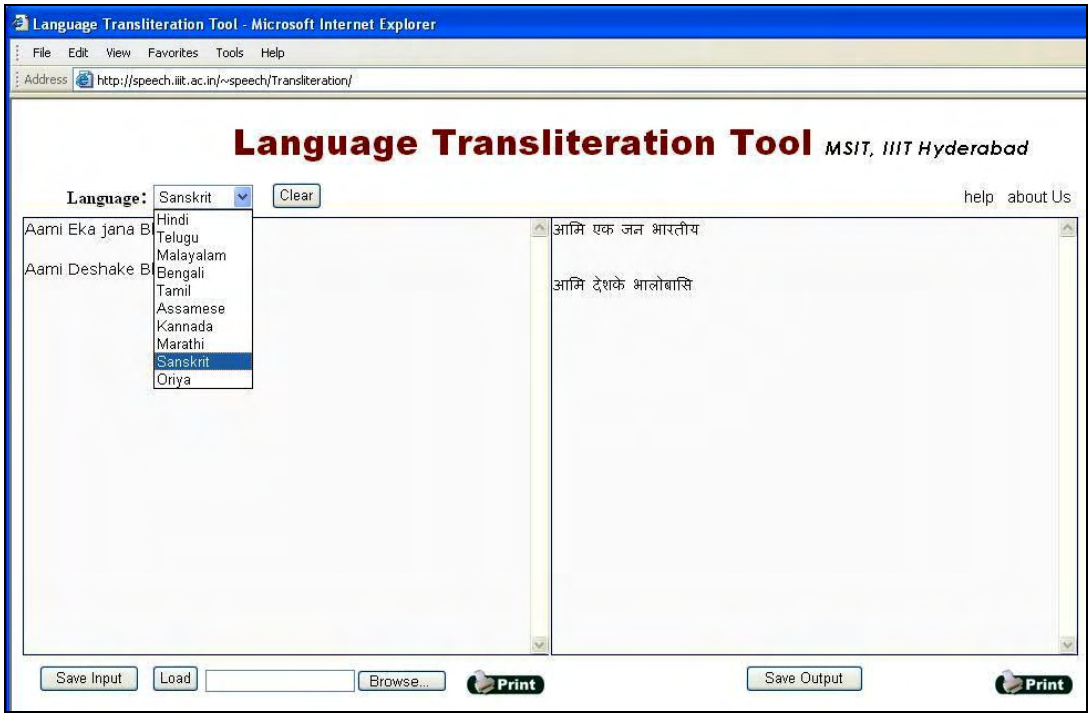


Fig. 2.46: Online Language Transliteration Tool for Indian Languages

2.13 Display of Search Results and Reading Books Onscreen

After getting user input in the online book search or browsing interface, DLI system efficiently retrieves data from the given parameters. DLI portal displays search results in the same window, where user earlier entered search strings or clicked on browsing pane. Search Result displays the brief details of retrieved books, alphabetically arranged with book title, along with metadata elements viz. name of author, barcode, year of publishing, language, subject and number pages. In a page at at time a list of 25 books is displayed with navigation tool to go to next few pages and last page. Search result also provides statistical detail relevant to the query, i.e., number of books matched along with total number of pages in those books.

When the user further selects a book and click on hyperlink, the Structural Metadata of selected book is displayed as shown in Figure 2.35. User can also view Admin Metadata of the book to know further details of book and scanning work as shown in Figure 2.36. Structural Metadata and Admin Metadata both provide hyperlinks to read book online. DLI Portal of RMSC Hyderabad provides a security check before opening a book reader. When user enters a code, given onscreen in a distorted image format, then he is allowed to proceed for reading (Figure 2.51). This unique feature, only available in IIIT Hyderabad hosted portal, probably is given for preventing systematic copying by crawlers.

User then read book online, one or two pages at a time depending on selection of BookReader-1 or BookReader-2. BookReader-1 displays one page at a time and BookReader-2 displays two consecutive pages at a time. For reading book online using BookReader, user has to download and install Alternatiff plugin once. This BookReader usually displays tiff images onscreen. Both online book readers provide standard onscreen image handling facilities such as zoom in, zoom out, etc. as available in any image viewer. For OCR-ed books, pages can also be displayed in HTML or TXT (text) or RTF format. These formats are usually available for books written in English or other European languages and went through OCR process.

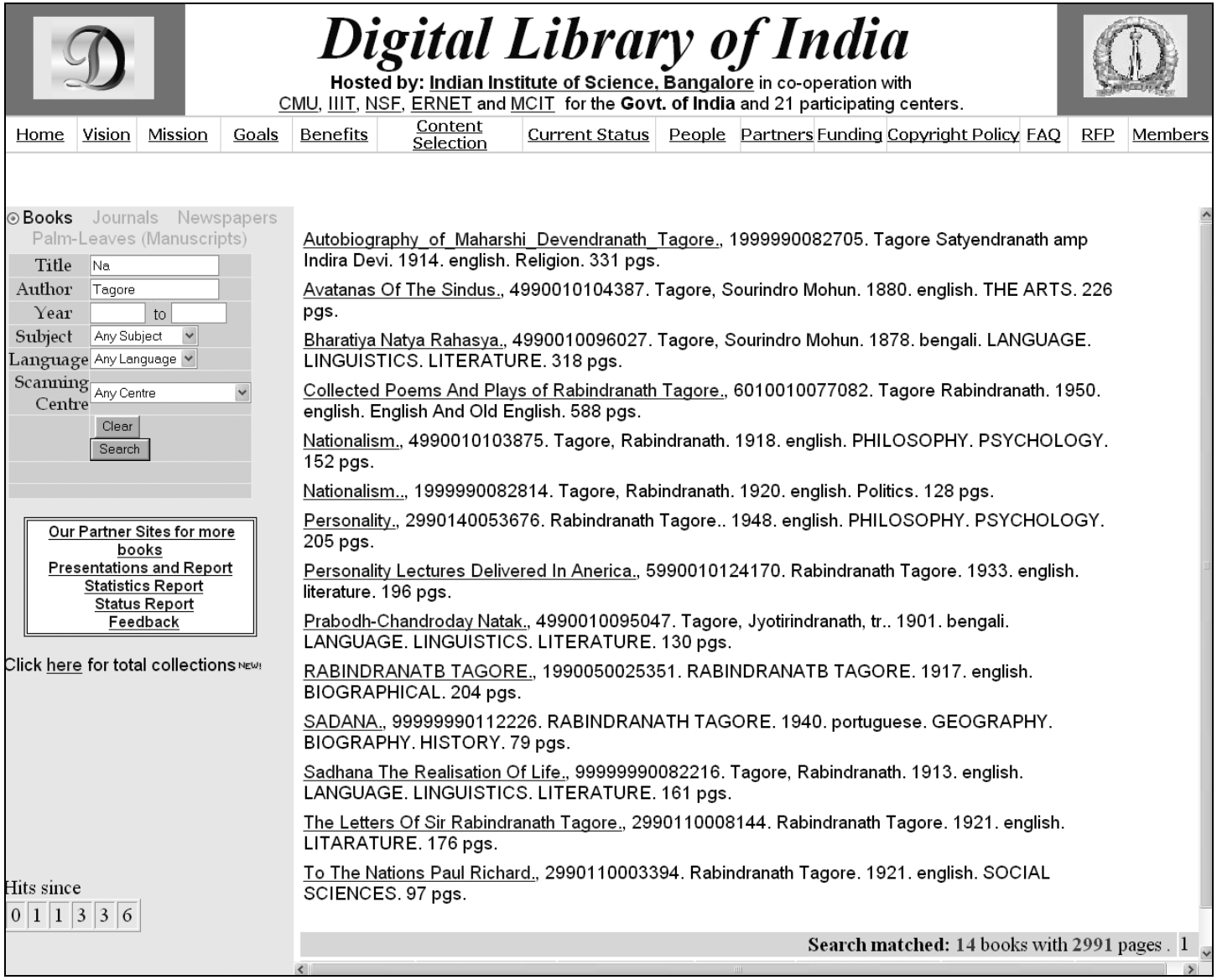


Fig. 2.47: Search Result Displayed in Main DLI Portal

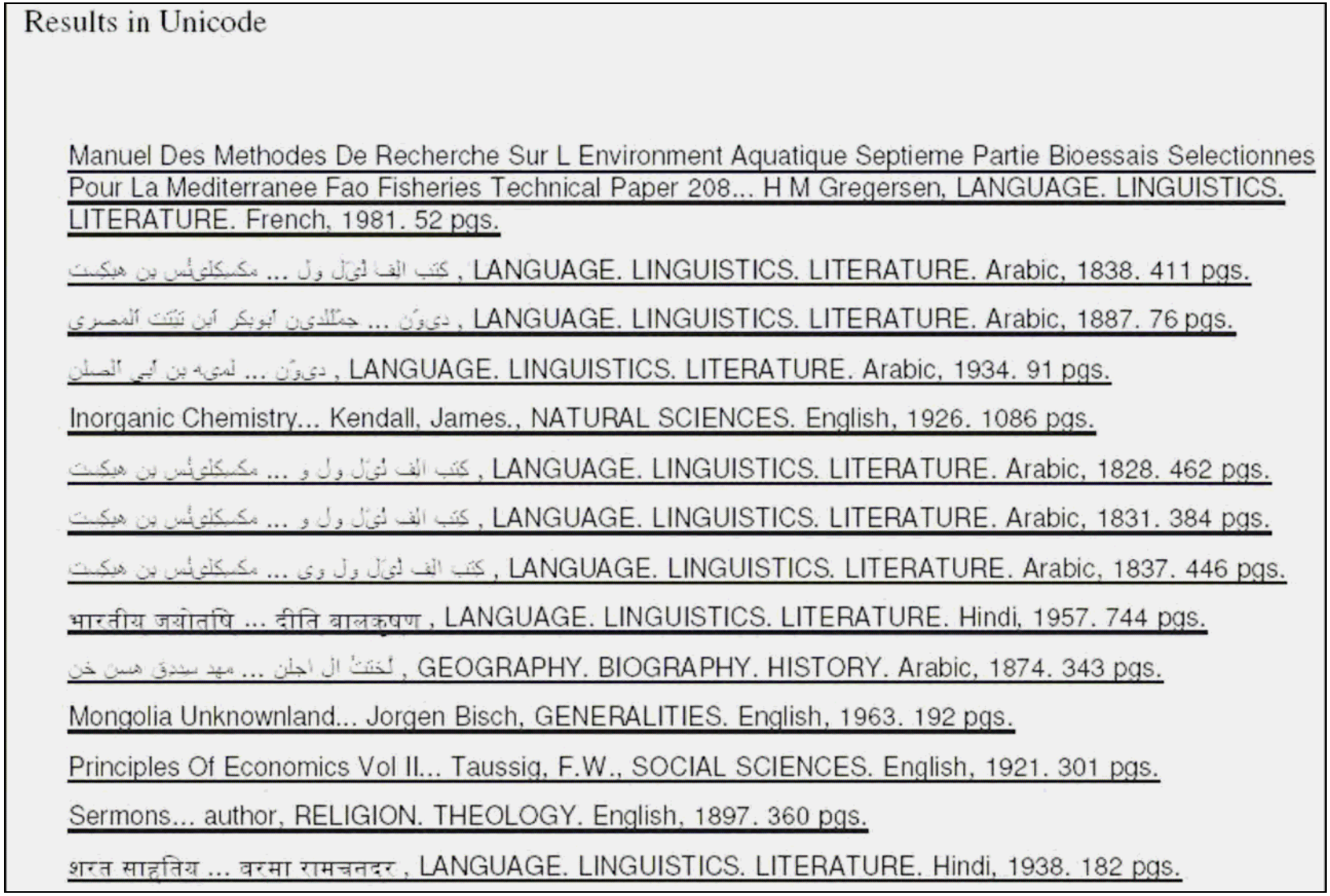


Fig. 2.48: Search Result Displayed in Unicode in DLI Portal of RMSC Hyderabad

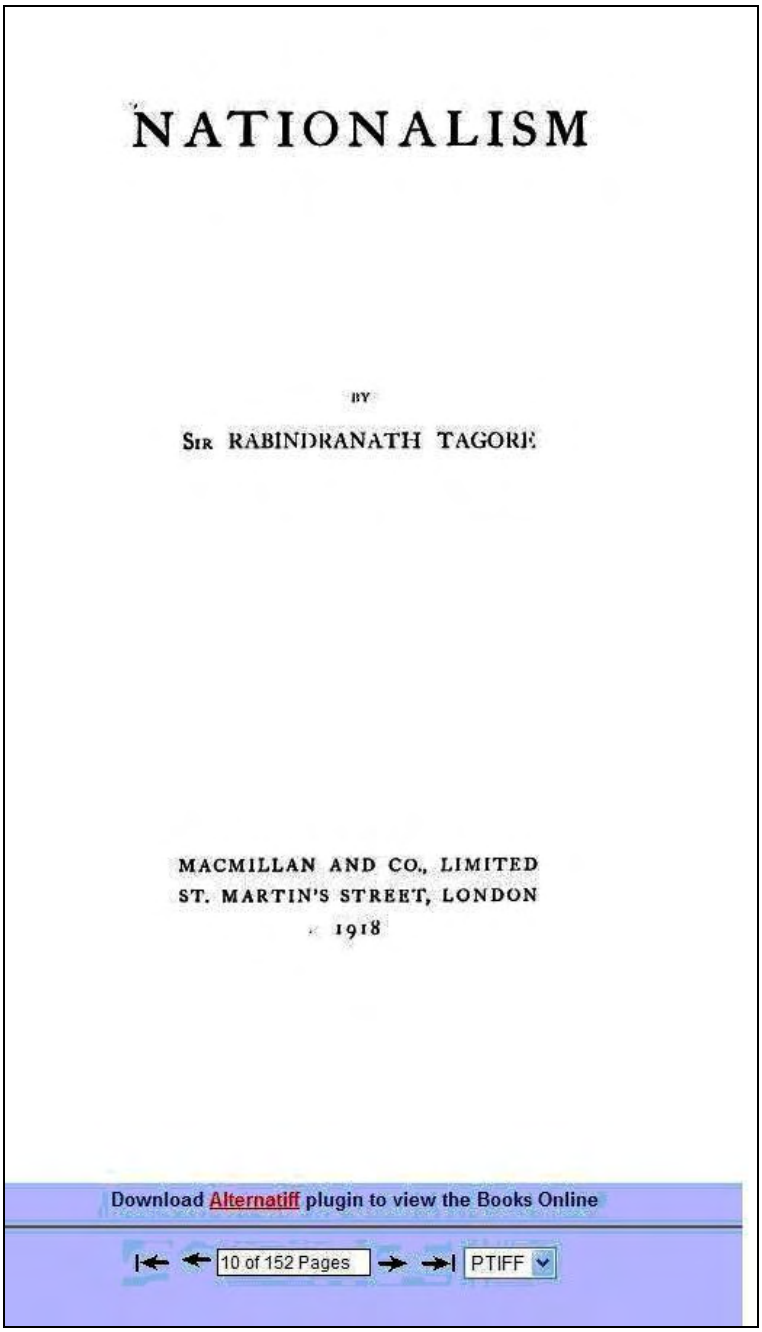


Fig. 2.49: BookReader-1: Book Displayed in TIFF Format – Single Page View

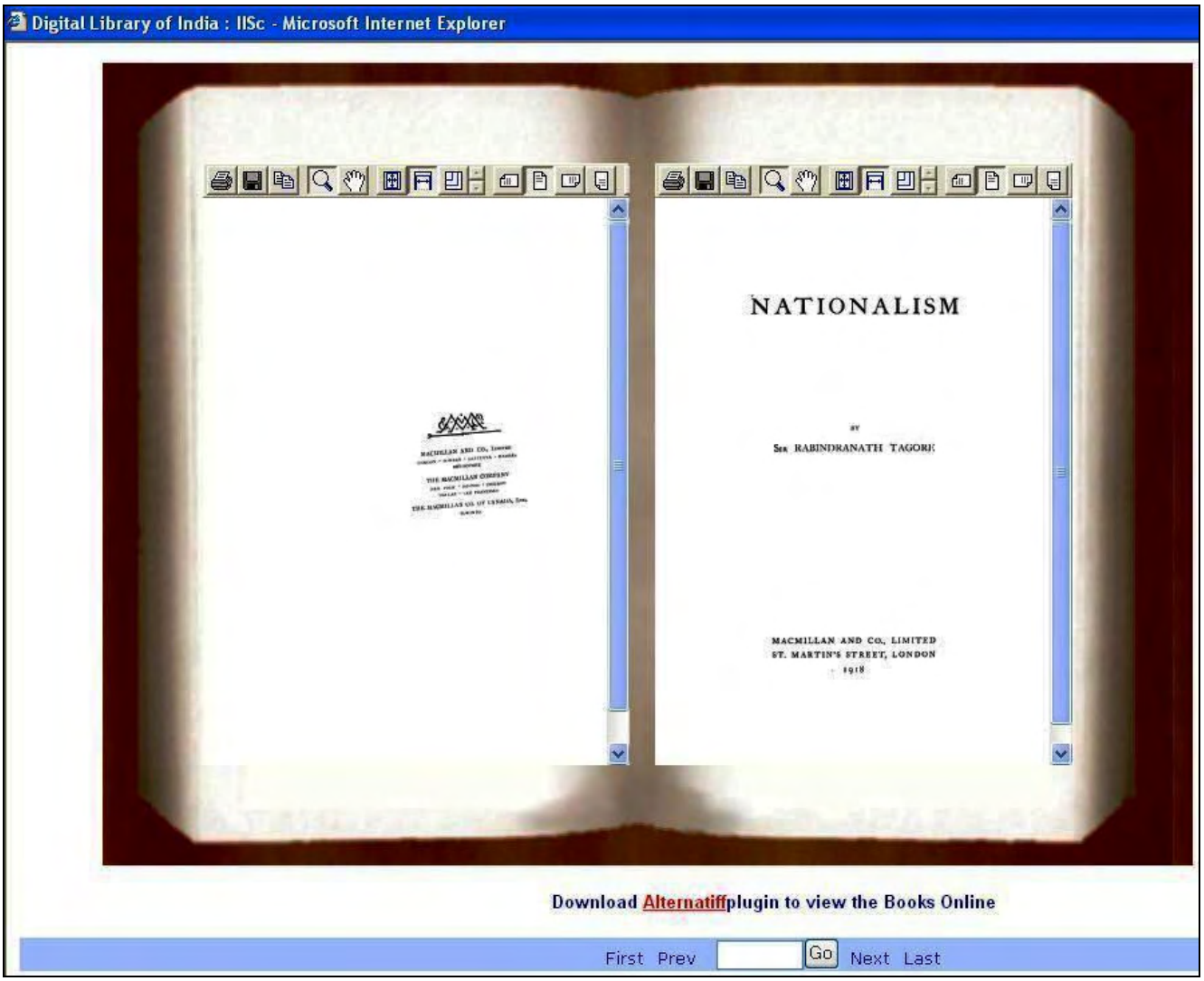


Fig. 2.50: BookReader-2: Book Displayed in TIFF Format – Double Pages View

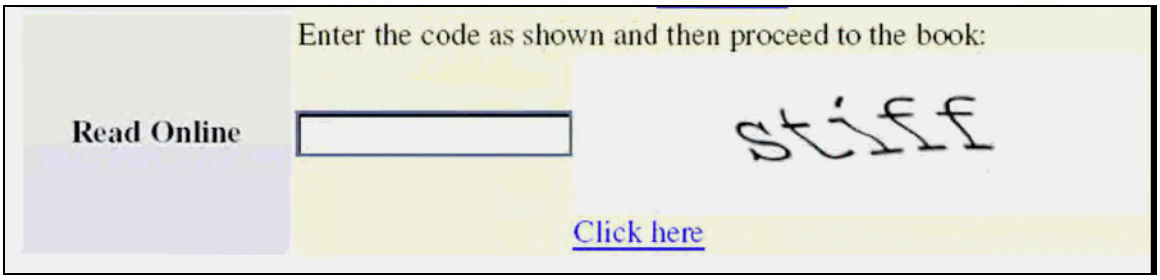


Fig. 2.51: Security Check for Reading Online in DLI Portal of RMSC Hyderabad

2.14 Statistical Analysis of Collections in DLI Portals

Language-wise Distribution of Digitized Books

Digital Library of India is a multilingual digital library initiative, where digitized books in different languages – both Indian and foreign – share the same digital library storage and retrieval system. Due to influence of foreign rulers in pre-independence India, English, Arabic, Persian, French and few other foreign languages became popular in many parts of the Indian subcontinent. India has 22 official languages, recognized by the Constitution of India. Hindi is the official language for Union and some state governments. English language also has official status. States of India also have their official languages. For example, Bengali is the official language in Tripura and West Bengal. Telugu is official language in Andhra Pradesh.

Indian publishers and institutions published millions of books in Indian languages since introduction of printing press in India in sixteenth century. Indian libraries, i.e., source libraries of DLI project, have developed significant collections of books in Indian and foreign languages, published in India and abroad. These significant collections are getting digitized in this effort and later being aggregated into DLI portals for providing worldwide access to literature available mainly in Indian libraries and institutions. Table 2.5 provides language-wise distribution of digitized books, available in main DLI portal. This Table shows only top ten languages, in terms of number of books available in DLI portal. This Table indicates that English stands at the top of the list and Telugu ranks second. Telugu is the topper amongst Indian languages.

Figure 2.52 graphically represents language-wise distribution of books in main DLI portal that indicates that English books comprise 63.9% of total digital library collection, Telugu books comprise a distant 11.4% and Hindi books 10.7%.

Figure 2.53 graphically represents language-wise distribution of scanned pages in main DLI portal that indicates English scanned pages comprise 69.7% of total available scanned pages in DLI portal, Hindi scanned pages comprise 9.4% and

Telugu scanned pages comprise a distant 7.4%. The percentage difference between the number of books and the number of scanned pages is only because of average number of pages in an English book is more than in any other Indian language book.

Figure 2.54 graphically represents language-wise distribution of books in DLI portal hosted by IIIT Hyderabad that indicates English books dominate this portal as well. English books comprise 54.3% of total digital library collection, whereas Telugu books comprise a distant 16.2% and Urdu books comprise 8.7%.

Figure 2.55 graphically represents language-wise distribution of scanned pages in DLI portal hosted by IIIT Hyderabad that indicates English scanned pages comprise 68.2% of total available scanned pages in DLI portal, whereas Telugu scanned pages comprise 7.7% and Urdu scanned pages comprise a distant 7.5%.

From Figures 2.54 and 2.55, we can assume that the dominating languages in the library collections in the state of Andhra Pradesh are English, Telugu, Arabic and Urdu.

Table 2.5 and Figures 2.52 to 2.55 also indicate some biasness in selection process of books for inclusion in DLI portal. Selection process is biased probably because of universal nature of DLI initiative that brings global audience to the DLI portals as a part of Universal Digital Library initiative, where naturally international language English has much more importance. On the other hand, books selected for digitization are mostly taken from public domain, i.e., out of copyrighted books or books published before 1924. Only a few Indian books can stand in that selection criteria. Still many books written in Indian languages are available in DLI portal, which are not even available in physical form in other states of India than originating state where source library is situated.

Table 2.5: Language-wise Distribution of Digitized Books

Rank	Language	Number of Books	Number of Pages
1	English	79,040	26,477,207
2	Telugu	14,090	2,812,929
3	Hindi	13,266	3,570,438
4	Sanskrit	4,854	1,820,761
5	Urdu	3,811	921,095
6	Tamil	1,617	411,723
7	Portuguese	1,220	138,021
8	Persian	1,186	305,769
9	Arabic	617	198,552
10	Bengali	554	338,213
	Others	3,356	1094,954
	Total	**123,611**	**38,089,662**
Language-wise Report as on 17 June 2008			

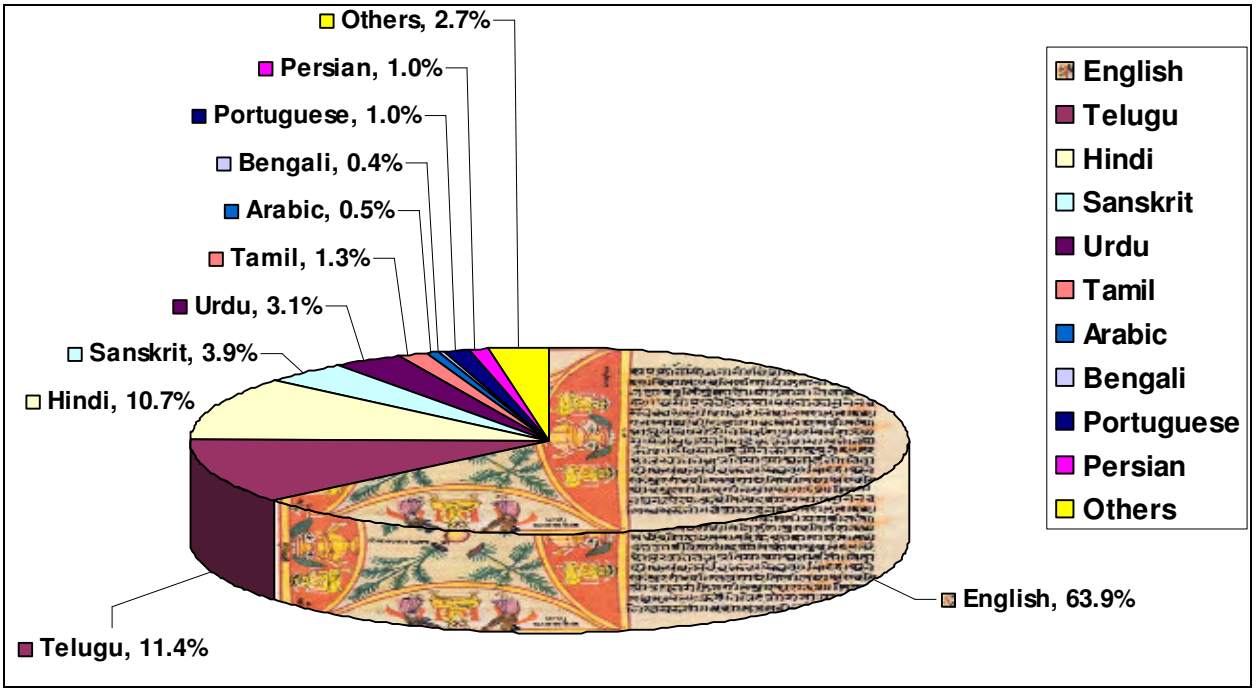


Fig. 2.52: Language-wise Distribution of Books in main DLI Portal

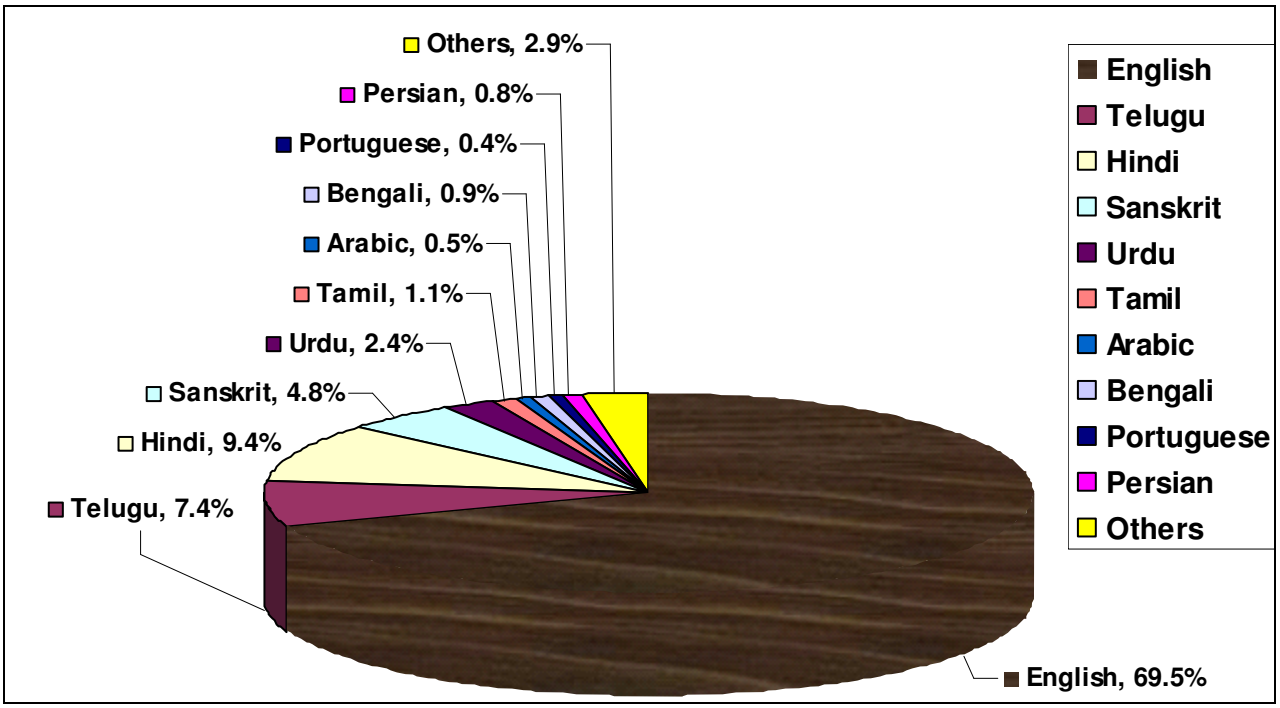


Fig. 2.53: Language-wise Distribution of Scanned Pages in Main DLI Portal

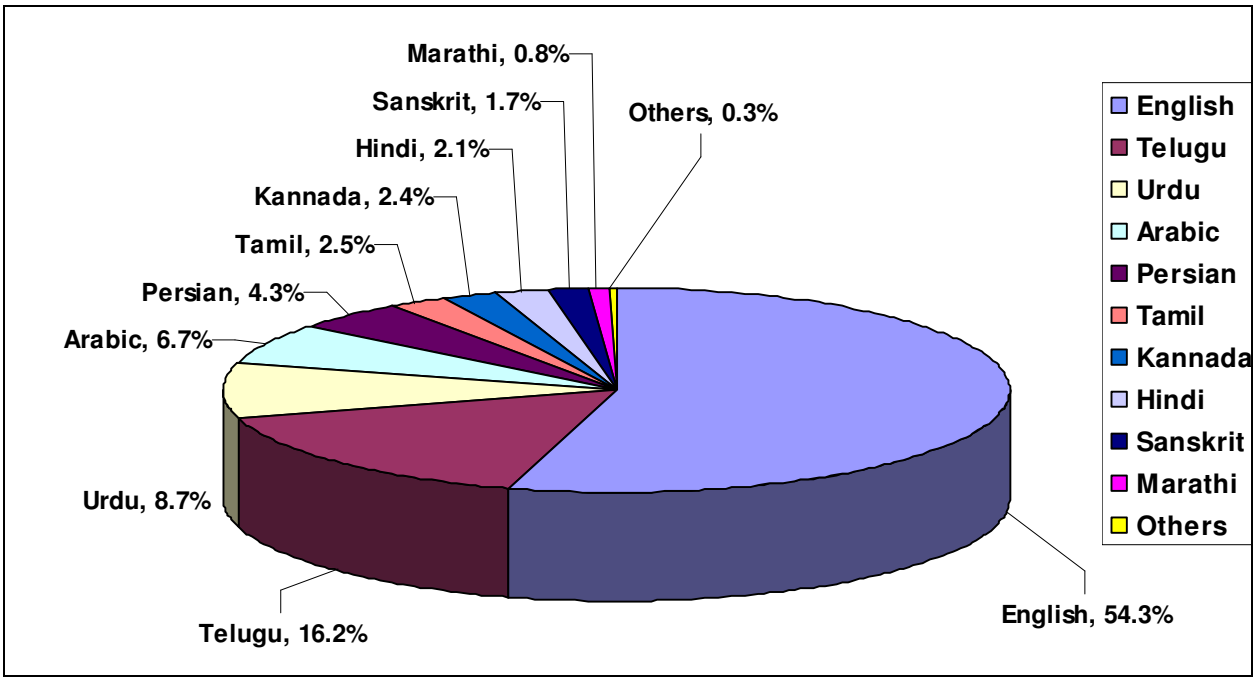


Fig. 2.54: Language-wise Distribution of Books in DLI Portal of RMSC Hyderabad

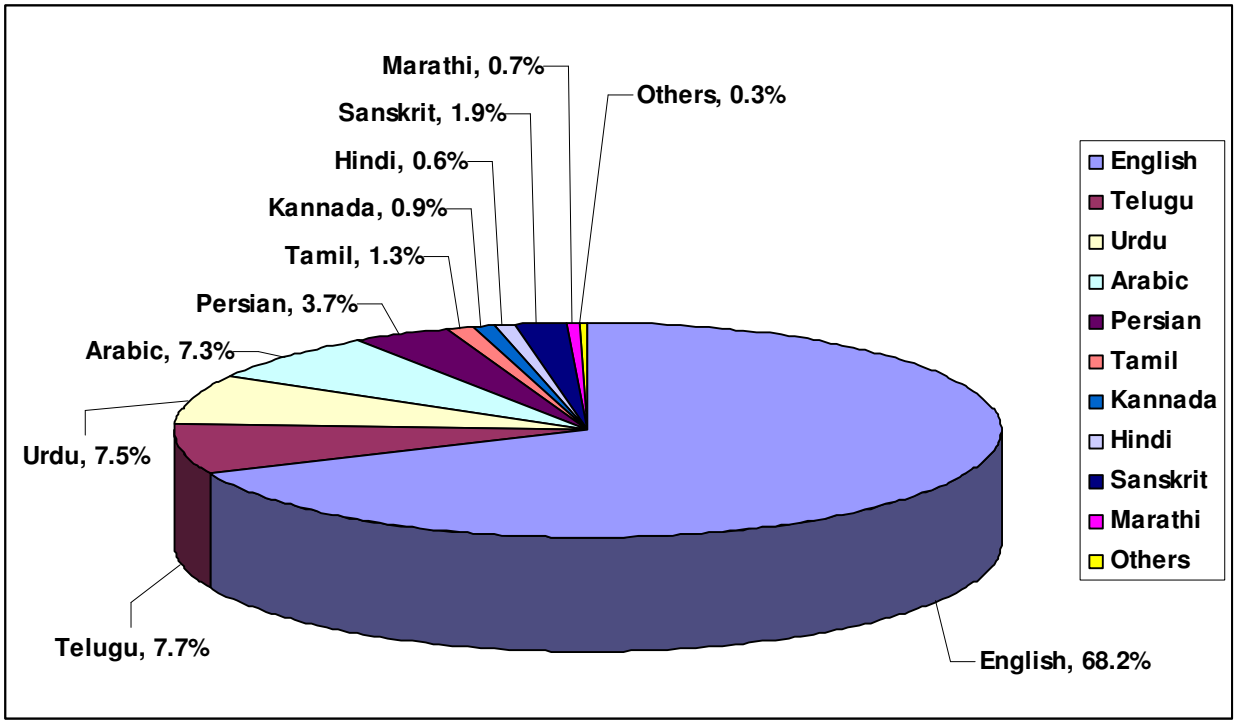


Fig. 2.55: Language-wise Distribution of Scanned Pages in DLI Portal of RMSC Hyderabad

Scanning Centre and Vendor--wise Distribution of Digitized Books

Digital Library of India project has country-wide geographical spread, where multi-cultural, multilingual diverse contents have been digitized and given open access to global audience. The scanning centres across the lengths and breadths of the country have digitized significant collections of wisdom, published in the book format.

Table 2.6 provides a ranked list of scanning centres across India, in terms of contributed digitized books in main DLI portal. This Table shows that Indian Institute of Information Technology (IIIT) Allahabad has contributed a significant number of digitized books and stands first in the ranked list. City Central Library in Hyderabad stands second and IIIT Hyderabad third in the list. This Table also shows that regional mega scanning centres, viz., IIIT Allahabad, IIIT Hyderabad and CDAC Kolkata, occupy the first, third and eleventh ranks respectively in terms of their contributions of digitized books to the main DLI portal.

Although IIIT Hyderabad hosts another DLI portal, it is still a contributing partner to the main DLI portal. Some scanning centres operating under this regional mega scanning centre such as City Central Library in Hyderabad, Tirumala Tirupati Devasthanams in Tirupati, State Central Library in Hyderabad, Salarjung Museum in Hyderabad, Osmania University in Hyderabad and Sri Venkateswara Oriental Research Institute in Tirupati are also contributing partners to the main DLI portal. On the other hand, regional mega scanning centre in CDAC Noida hosts another DLI portal, but it is not a contributing partner to the main DLI portal. Two scanning centres operating under regional mega scanning centre at CDAC Noida, viz. Rashtrapati Bhavan and Bharatiya Jnanpith, are also contributing content partners of main DLI portal. Figure 2.56 depicts the scanning-centre-wise distribution of digitized books in main DLI portal.

CDAC Kolkata hosts regional mega scanning centre for eastern region. It does not have its own DLI portal. It provides digitized contents to main DLI portal. The search interface of DLI portal provides searchability within the collection of a scanning centre. Figure 2.57 depicts language-wise distribution of books, digitized under regional mega scanning centre in Kolkata as on 30th June 2008. This Figure indicates that English books have been predominantly digitized in this regional mega scanning centre (70.34%), followed by Bengali books (12.54%). Almost all Bengali books available in main DLI portal have been digitized by the regional mega scanning centre in Kolkata.

Figure 2.58 depicts subject-wise distribution of books, digitized under regional mega scanning centre in Kolkata as on 30th June 2008. This Figure indicates that top three subject areas of scanned books are History, Geography and Science. Some digitized books on religion and literature have been uploaded into main DLI portal from this regional mega scanning centre as well.

Table 2.7 provides a ranked list of scanning centres operated under regional mega scanning centre at IIIT Hyderabad, in terms of contributed digitized books in DLI portal hosted by IIIT Hyderabad. This Table shows that IIIT Hyderabad itself is the major contributor of digitized contents into DLI portal and stands first

in the ranked list. State Central Library in Hyderabad stands second and City Central Library in Hyderabad third in the list. Figure 2.59 depicts the scanning-centre-wise distribution of digitized books in DLI portal hosted by regional mega scanning centre in Hyderabad.

In regional mega scanning centre at IIIT Hyderabad, three scanning vendors provide digitization services to the DLI project. The vendors are responsible for execution of scanning work in mutually agreed timeframe without compromising quality of work. These vendors represent DLI partners from the small and medium sized enterprise (SME) segment.

Other regional mega scanning centres and scanning centres across India also utilize digitization services from the third party vendors. But DLI portal hosted in Hyderabad only provides access point to search by the name of the vendor.

Table 2.8 provides a ranked list of vendors operated under regional mega scanning centre at IIIT Hyderabad, in terms of contributed digitized books in DLI portal hosted by IIIT Hyderabad. This Table shows that Par Informatics Limited, a Hyderabad-based vendor is the major contributor to digitized contents into DLI portal and stands first in the ranked list. Thrinaina Informatics Limited in Hyderabad stands second and SV Infosys in Tirupati third in the list. SV Infosys is affiliated to Tirumala Tirupati Devasthanams in Tirupati. Figure 2.60 depicts the vendor-wise distribution of digitized books in DLI portal hosted by regional mega scanning centre in Hyderabad.

Table 2.6: Scanning Centre-wise Distribution of Digitized Books

Rank	Scanning Centre	Number of Books	Number of Pages	Number of Pages In Millions
1	IIIT Allahabad	37,633	10,669,514	10.67
2	City Central Library, Hyderabad	12,952	2,762,568	2.763
3	IIIT Hyderabad	11,759	5,121,265	5.121

4	Sri Venkateswara Digital Library, Tirumala Tirupati Devasthanams, Tirupati	11,705	3,019,386	3.019
5	AKCE, Virudhunagar	7095	1,036,014	1.036
6	State Central Library, Hyderabad	6479	2,281,804	2.282
7	Salarjung Museum, Hyderabad	6128	2,339,384	2.339
8	Osmania University, Hyderabad	5779	1,975,659	1.976
9	Rashtrapati Bhavan Library, New Delhi	4832	2,023,364	2.023
10	IISc Bangalore	4418	1,620,729	1.621
11	CDAC Kolkata	4353	2,056,937	2.057
12	SASTRA, Thanjavur	1872	577,050	0.577
13	University of Pune	1686	497,074	0.497
14	Indian Institute of Astrophysics, Bangalore	1568	560,577	0.561
15	Academy of Sanskrit Research, Melkote	1297	394,001	0.394
16	Goa University	1214	134,481	0.134
17	Sringeri Sharada Peetham	958	306,439	0.306
18	Sri Venkateswara Oriental Research Institute, Tirupati	495	216,271	0.216
19	Digital Library, Mulund	492	160,185	0.16
20	Bharatiya Jnanpith, Delhi	292	92,570	0.093
21	Anna University, Chennai	247	122,904	0.123
	Others	357	121,486	
	Total	***123,611***	***38,089,662***	***38.089***
Scanning Centre-wise Report as on 17 June 2008				

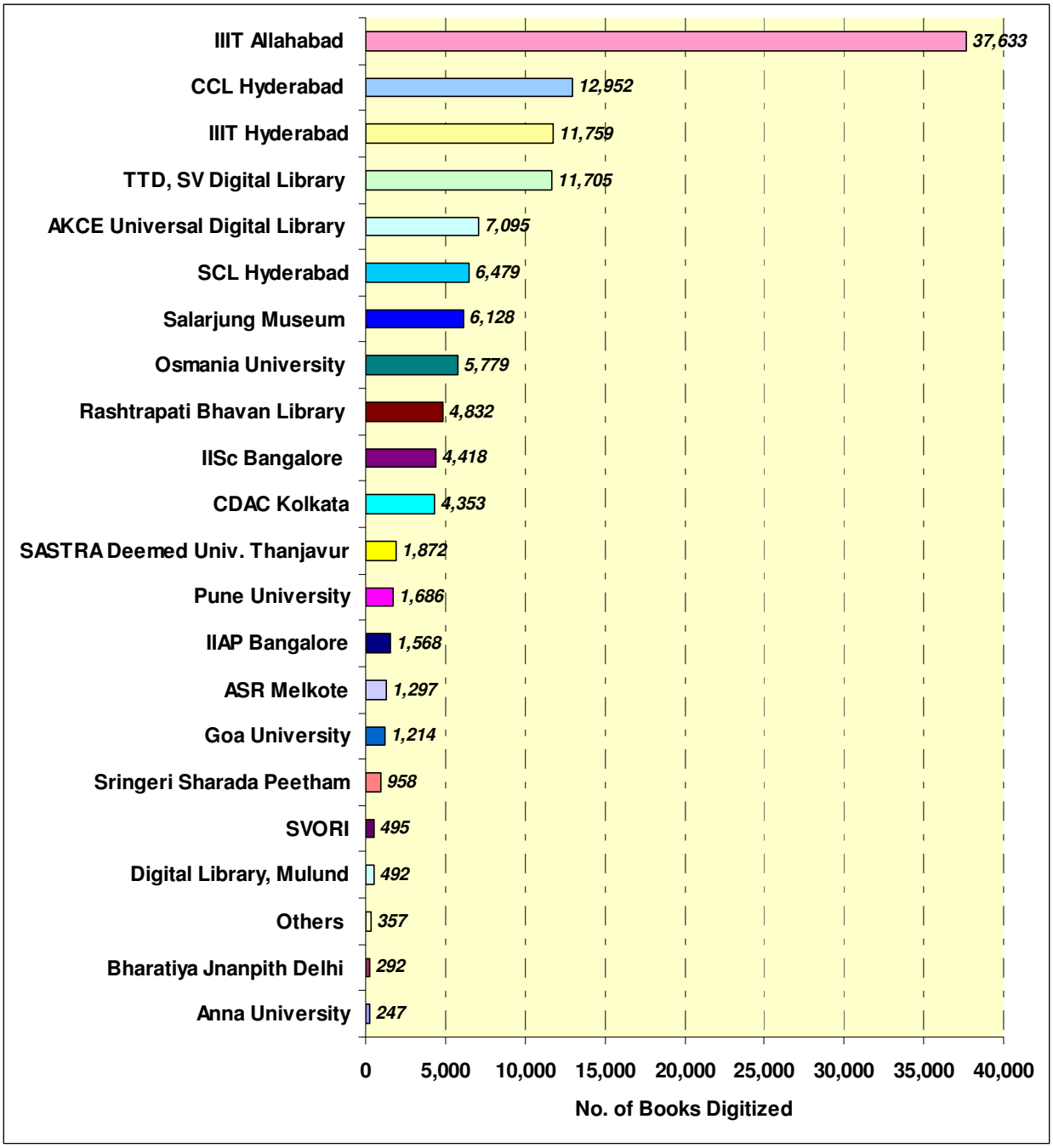


Fig. 2.56: Scanning Centre-wise Distribution of Digitized Books

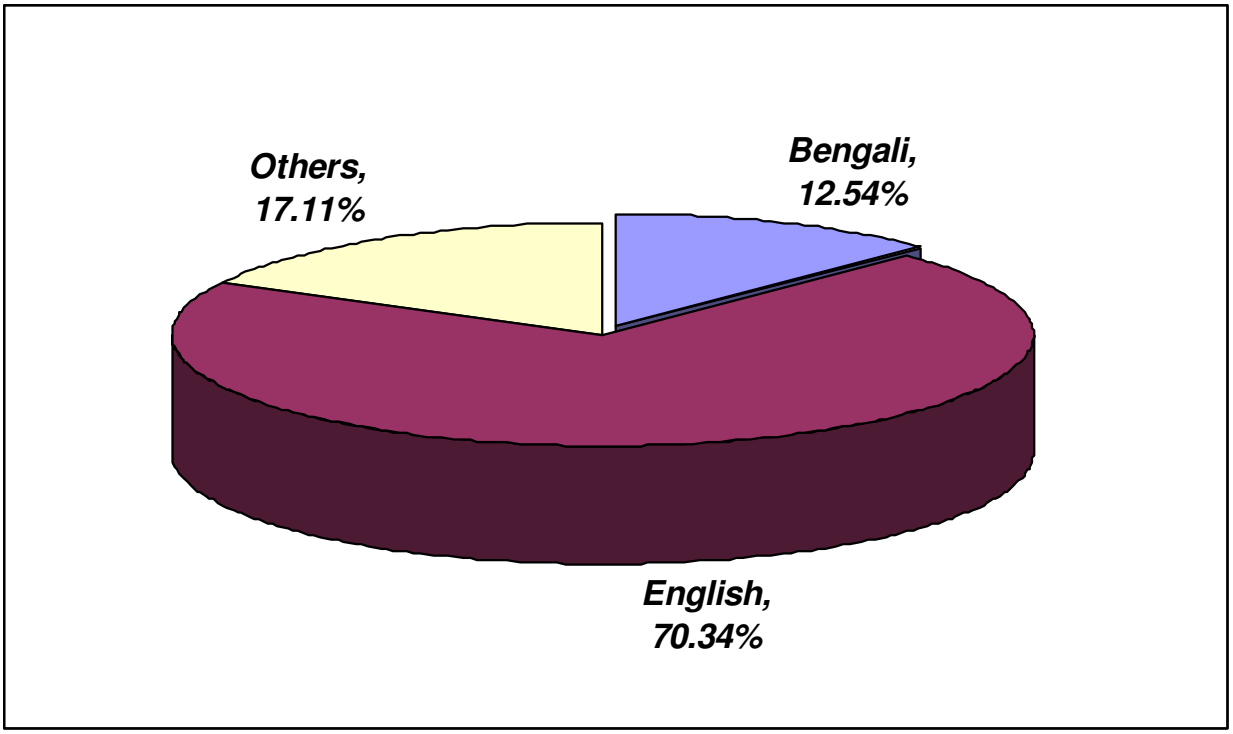


Fig. 2.57: Language-wise Distribution of Books, Digitized in RMSC Kolkata

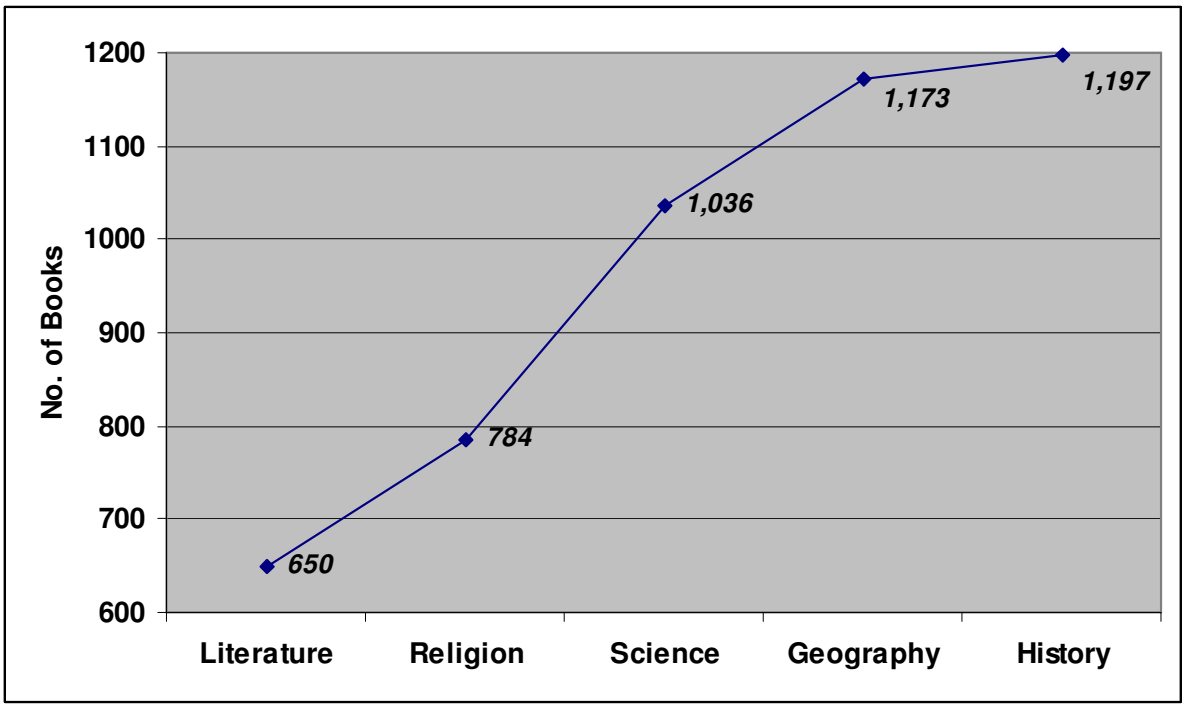


Fig. 2.58: Subject-wise Distribution of Books, Digitized in RMSC Kolkata

Table 2.7: Scanning Centre-wise Distribution of Digitized Books, operated under RMSC Hyderabad

Rank	Name of Scanning Centre	Number of Books	Number of Pages	Number of Pages in Millions
1	Regional Mega Scanning Centre, IIIT Hyderabad	27,893	8,993,360	8.993
2	State Central Library (SCL), Andhra Pradesh	4,489	1,463,897	1.464
3	City Central Library (CCL), Hyderabad	1,436	412,900	0.413
4	Sri Venkateswara Digital Library, Tirumala Tirupati Devasthanams (TTD), Tirupati	1,318	288,429	0.288
5	Hyderabad University (HCU)	14	6,567	0.007
	Others	25,396	8,009,367	8.009
	Total	**60,546**	**19,174,520**	**19.175**
As on 1 July 2008				

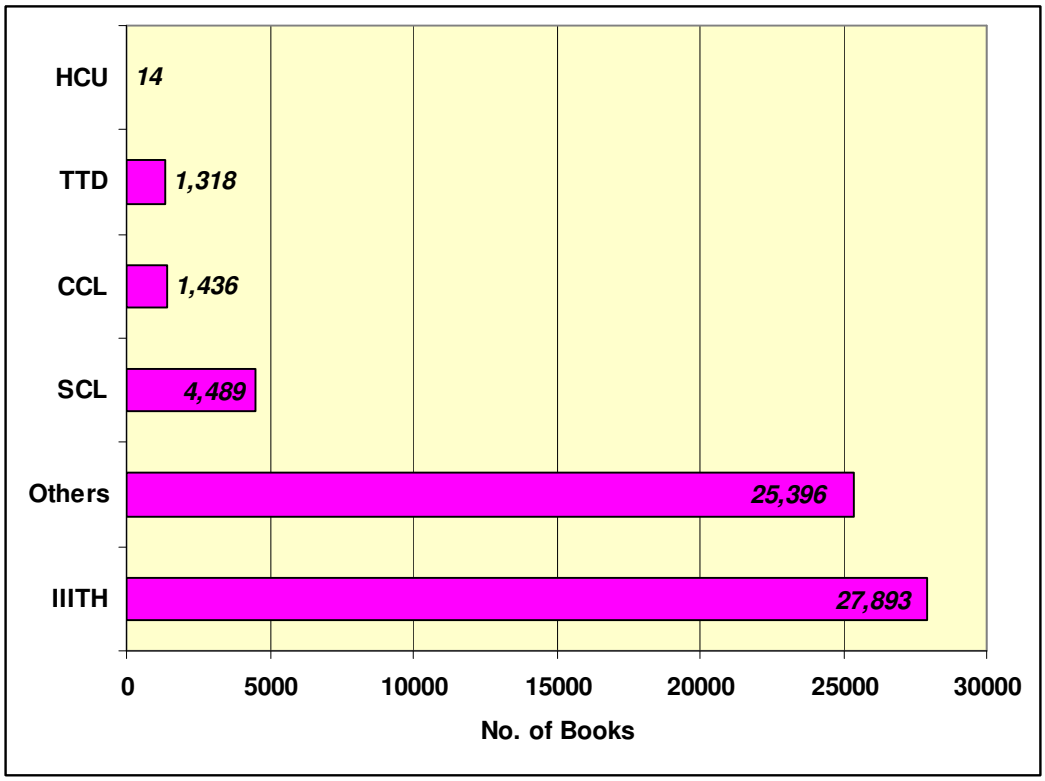


Fig. 2.59: Scanning Centre-wise Distribution of Digitized Books, operated under RMSC Hyderabad

Table 2.8: Vendor-wise Distribution of Digitized Books, operated under RMSC Hyderabad

Rank	Name of Vendor	Number of Books	Number of Pages
1	Par Informatics Limited, Hyderabad, Andhra Pradesh (PAR)	19,814	6,394,118
2	Thrinaina Informatics Limited, Hyderabad, Andhra Pradesh (TIL)	9,885	3,709,802
3	SV Infosys, Tirumala Tirupati Devasthanams, Tirupati, Andhra Pradesh (SVI)	5,904	1,085,890
As on 1 July 2008			

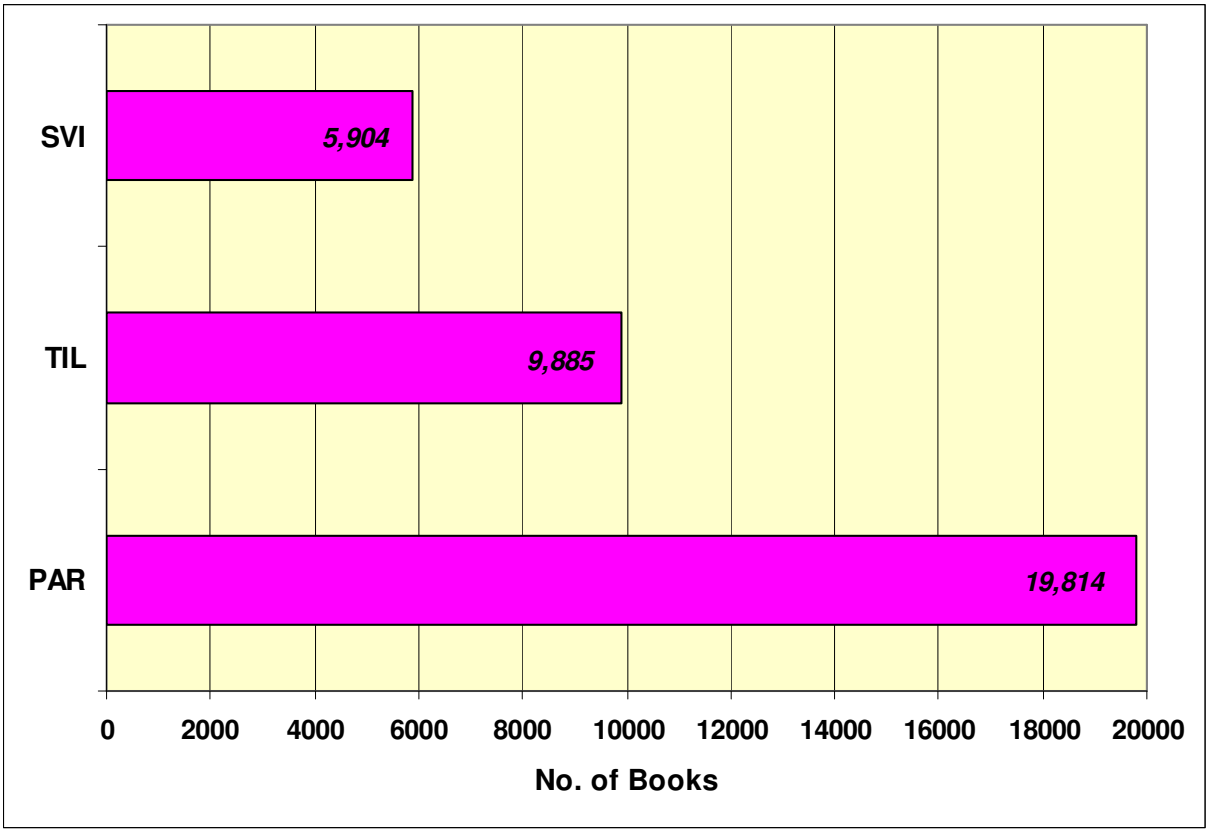


Fig. 2.60: Vendor-wise Distribution of Digitized Books, operated under RMSC Hyderabad

Subject-wise Distribution of Digitized Books

In any information storage and retrieval system, subject is an important point of access to reach the relevant documents of a user's choice. DLI system essentially provides easy access to subject categories both from the navigation window panel and search panel. In main DLI portal, 34 subject areas have been identified and can be searched/ navigated by any of the subject area.

Table 2.9 provides a ranked list of subject-wise distribution of digitized books hosted in main DLI portal. The subject category Literature secures first rank, followed by Science (second) and History (third). This list identifies other top ten subject categories also which are: Religion, Geography, Philosophy, Psychology, General, Art, and Engineering.

This Table also indicates that more than 52,000 books have unassigned subject categories. This is a major limitation in DLI project, because without having a subject category, these books are unlikely to be retrieved and accessed by end users.

Sometimes, for a book, more than one subject category is assigned, when the book is multidisciplinary or interdisciplinary in nature. In that case, some books have more than one assigned subject category whereas some other books have unassigned subject categories due to inattention in metadata quality checking process.

Figures 2.61 and 2.62 depict the subject-wise distribution of digitized books in main DLI portal and DLI portal hosted by IIIT Hyderabad respectively.

Table 2.10 provides a ranked list of subject-wise distribution of digitized books in DLI portal hosted by IIIT Hyderabad. Here also, subject category Literature secures first rank, followed by History (second) and Religion (third). This list identifies other top ten subject categories, which are: General, Education, Philosophy, Art, Poetry, Chemistry and Sociology. In this DLI portal books are not available in some subject categories such as Astronomy and Commerce.

Table 2.9: Subject-wise Distribution of Digitized Books

Rank	Subject	Number of Books	Number of Pages	No of Pages In Millions (approx)
1	Literature	27,277	9,066,027	9.066
2	Science	12,068	4,041,056	4.041
3	History	7,158	2,381,669	2.382
4	Religion	5,865	1,981,556	1.982
5	Geography	5,002	1,535,218	1.535
6	Philosophy	4,038	1,359,908	1.360
7	Psychology	3,377	1,186,654	1.187
8	General	1,855	740,196	0.740
9	Art	771	220,152	0.220
10	Engineering	676	104,491	0.104
11	Chemistry	365	104,640	0.105
12	Law	313	164,227	0.164
13	Poetry	313	50,637	0.051
14	Economics	289	91,557	0.092
15	Mathematics	245	78,030	0.078
16	Health	244	31,068	0.031
17	Electrical	205	31,202	0.031
18	Astronomy	203	80,802	0.081
19	Education	172	23,251	0.023
20	Biology	126	44,435	0.044
21	Music	126	36,093	0.036
22	Medical	121	30,471	0.030
23	Astrophysics	103	38,613	0.039
24	Politics	87	30,882	0.031
25	Sociology	86	27,136	0.027
26	Marketing	80	3,324	0.003
27	Statistics	73	16,021	0.016

28	Devotional	35	5,759	0.006
29	Autobiography	26	9,900	0.010
30	Architecture	24	9,274	0.009
31	Zoology	12	3,655	0.004
32	Commerce	9	2,510	0.003
33	Mythology	6	887	0.001
34	Dance	3	515	0.001
35	Unknown	2	513	0.001
36	No Subject Metadata Available	52,320	14,565,342	14.565
	Total	**123,675**	**38,097,671**	**38.098**
Subject-wise Report as on 26 June 2008				

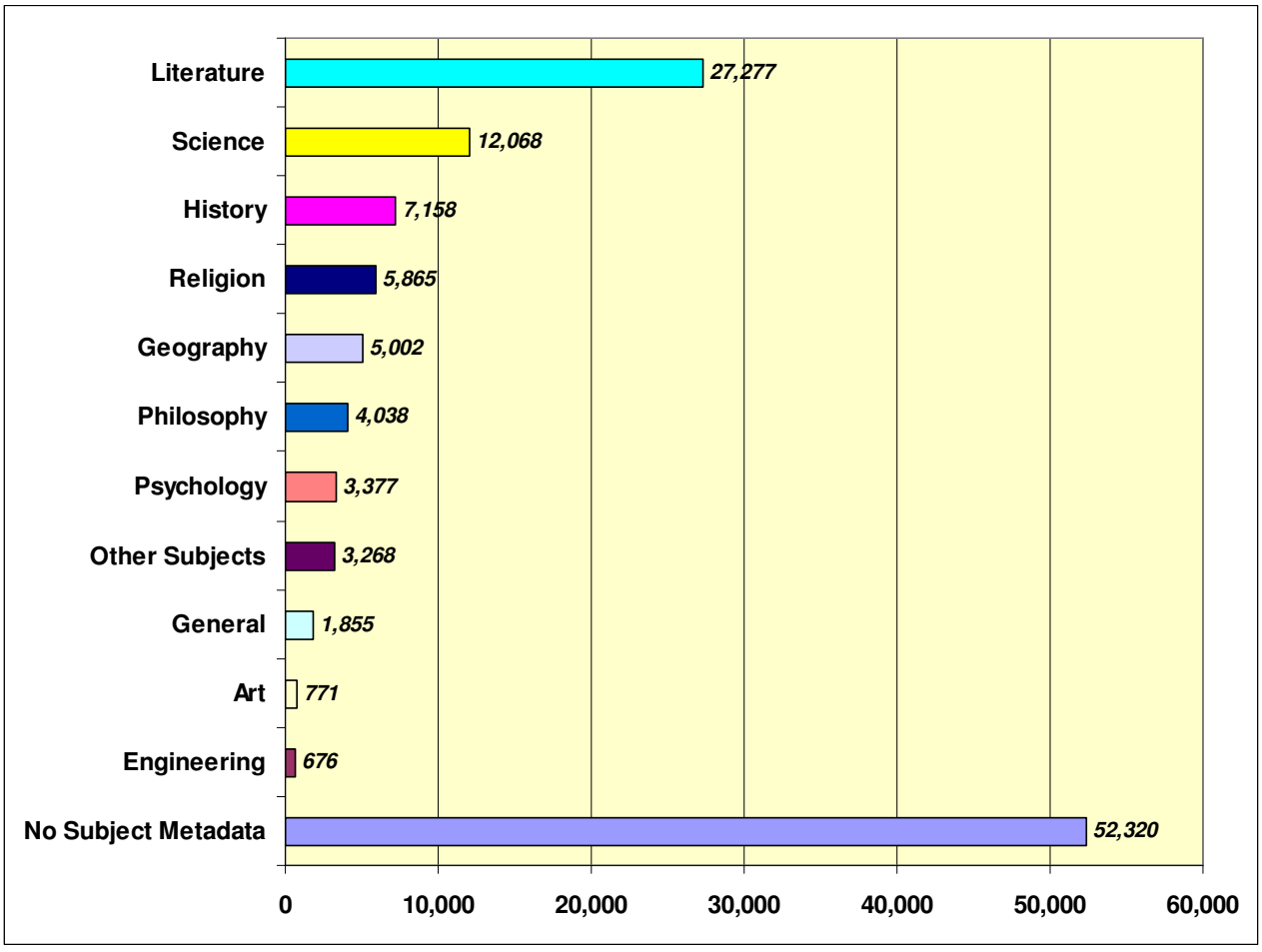


Fig. 2.61: Subject-wise Distribution of Digitized Books

Table 2.10: Subject-wise Distribution of Digitized Books in DLI Portal of RMSC Hyderabad

Rank	Subject	Number of Books	Number of Pages	No of Pages in Million (approx)
1	Literature	3,291	889,463	0.889
2	History	3,276	1,749,435	1.749
3	Religion	1,962	477,262	0.477
4	General	1,452	454,085	0.454
5	Education	356	155,335	0.155
6	Philosophy	349	100,875	0.101
7	Art	337	106,995	0.107
8	Poetry	295	79,405	0.079
9	Chemistry	272	98,052	0.098
10	Sociology	232	84,780	0.085
11	Economics	216	71,085	0.071
12	Mathematics	244	62,842	0.063
13	Psychology	193	72,501	0.073
14	Biology	139	53,669	0.054
15	Science	133	40,750	0.041
16	Politics	126	47,364	0.047
17	Geography	112	30,564	0.031
18	Music	107	34,071	0.034
19	Medical	80	40,766	0.044
20	Law	74	29,996	0.041
21	Engineering	64	21,992	0.022
22	Astronomy	48	19,344	0.019
23	Marketing	38	12,411	0.012
24	Zoology	35	13,720	0.014
25	Architecture	31	12,012	0.012
26	Mythology	25	11,679	0.012

27	Autobiography	18	3,727	0.004
28	Statistics	16	4,381	0.004
	Unknown	2,187	670,697	0.671
	No Subject Metadata Available	44,999	13,736,422	13.736
	Total	**60,707**	**19,185,680**	**19.186**
Subject-wise Report as on 19 July 2008				

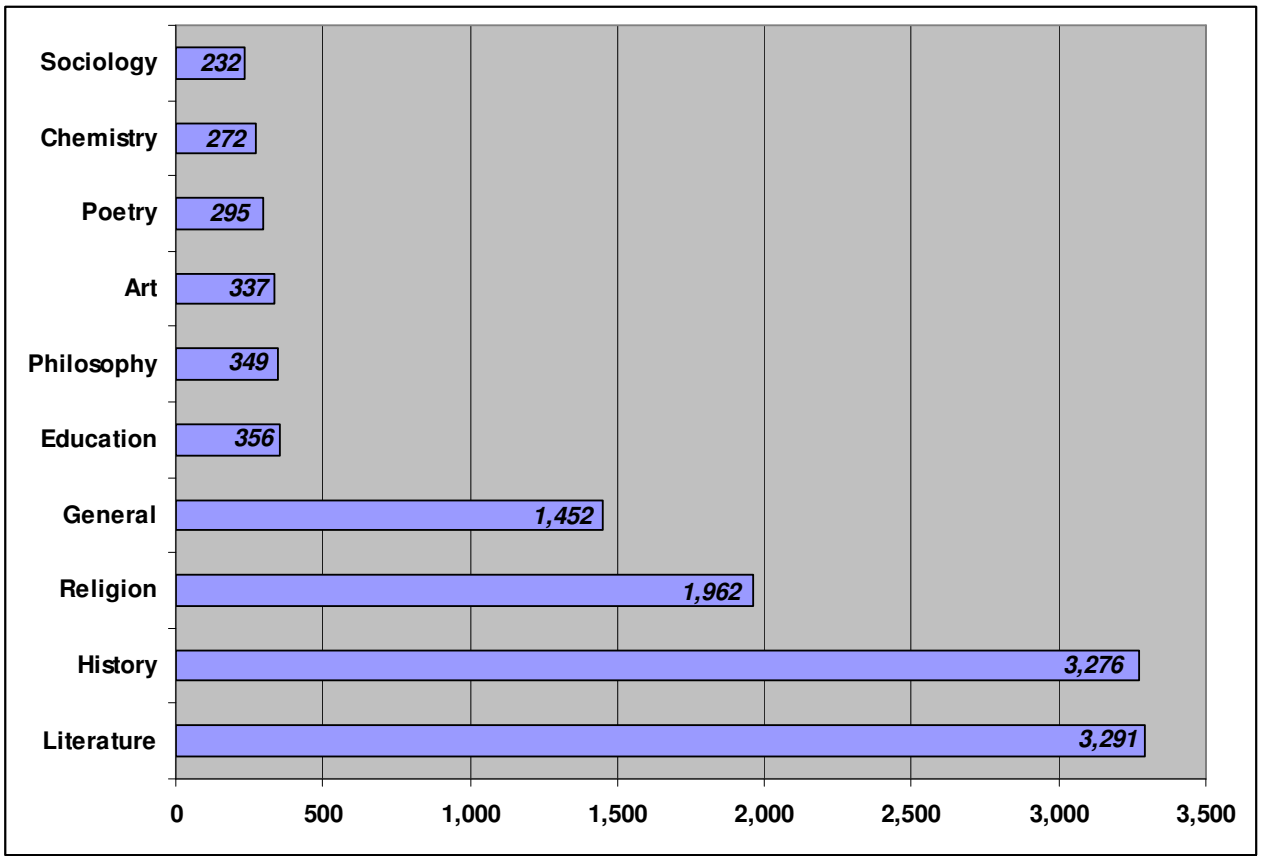


Fig. 2.62: Subject-wise Distribution of Digitized Books in DLI Portal of RMSC Hyderabad (Top Ten Subject Categories)

2.15 Architecture of DLI Portals

Digital Library of India initiative maintains three Web-portals for providing open access to digital library contents. Indian Institute of Science Bangalore (IISc Bangalore) is the national coordinator of this project and hosts main DLI portal. Regional mega scanning centres at the International Institute of Information Technology Hyderabad (IIIT Hyderabad) and Centre for Development of Advanced Computing Noida (CDAC Noida) host two more DLI portals in their respective locations. Other two regional mega scanning centres at the Indian Institute of Information Technology Allahabad (IIIT Allahabad) and Centre for Development of Advanced Computing Kolkata (CDAC Kolkata) only produce digital contents and supply these contents to the DLI national coordination office at IISc Bangalore for uploading into main DLI portal.

A DLI portal usually maintains at least two main servers, one as a metadata server and another as data server. In metadata server, metadata elements of all books available with this portal are maintained. Data servers are usually terabyte servers, responsible for hosting digitized contents in all supported formats such as TIFF (image), TXT (text), RFT (formatted text) and HTML (webpage). A robust local area network is established in the regional mega scanning centre, which is hosting a DLI portal. This network enables interactive communication between servers for fetching relevant metadata, scanned pages and texts. Some proxy servers are also established in the RMSC that accept requests from clients and forward them to metadata and data servers. Proxy servers also improve network operations and security, and act as a firewall, mediating traffic between the protected internal DLI network and the Internet.

Figure 2.63 depicts the network architecture in a regional mega scanning centre that hosts a DLI portal. This Figure identifies linkages between the external traffic, i.e., client computers and internal local area network. For reducing vulnerability of network security, firewall is established between Internet and proxy servers in RMSC.

The data servers in DLI site have series of servers to store the digital images of pages from scanned books. Images in TIFF format require a considerable large storage space, because these are uncompressed loss-less image files. As DLI maintains other file formats such as TXT (text), RFT (formatted text) and HTML (webpage), data servers are integrated for storage of all supported file formats. The hyperlinks from metadata servers to data servers should be routinely verified in regular interval to address the problem of hyperlink failures, i.e., reducing 'page not found' notices.

Digital Library of India portals try to address Web accessibility issue, particularly making DLI portals usable by people of all abilities and disabilities. A new architecture is proposed to make contents of the portals audible to the visually impaired persons, by introducing a parallel Text-to-Speech (TTS) server. Visual impairments include blindness, low vision, poor eyesight and color blindness. TTS refers to ability of the computer to convert text into spoken words. The TTS server, with the help of TTS software tools, will store and disseminate converted spoken words from DLI document repositories to the visually impaired end users. Figure 2.64 provides a simple data flow diagram for a DLI portal integrated with Text-to-Speech (TTS) System.

So far DLI network architecture has proved to be as an efficient model, with self-healing mechanisms to address the network failure. If any of the nodes of DLI network temporarily breaks down, this architecture allows DLI portal operational without reducing any efficiency (Ambati & Reddy, 2005).

All DLI portals maintain similar kind of architecture, only number of data servers varies depending upon the collection size.

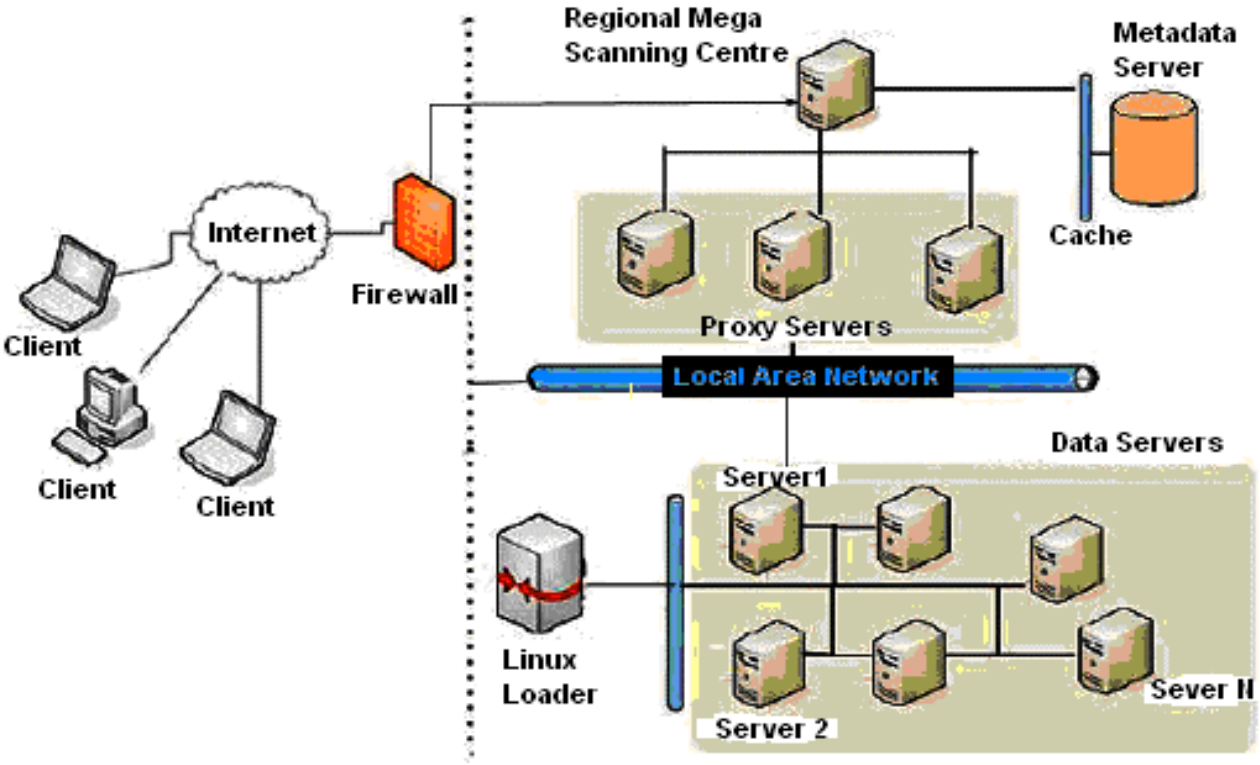


Fig. 2.63: Network Architecture in Regional Mega Scanning Centre that Hosts a DLI Portal

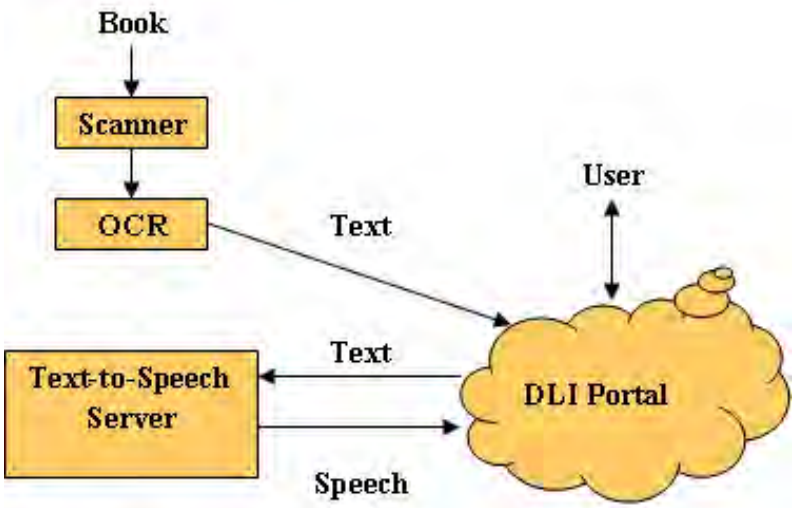


Fig. 2.64: Data Flow in a DLI Portal integrated with Text to Speech (TTS) System

2.16 Benefits of DLI Initiative at the National Scale

Digital Library of India initiative provides many opportunities to all segments of learners and readers, including persons located in remote places but having access to ICT facilities at their neighbourhoods. Some of the benefits have already been achieved by the DLI. The benefits achieved are described below as well as indicated in Figure 2.65:

- Supplements the formal education system by making knowledge available to anyone who can read and has access;
- Makes available the out-of-print, out-of-circulation, out-of-copyright and un-copyrighted books, magazines, reports and other useful documents, originated from Indian subcontinent;
- Makes available the classic literature – originated from Indian subcontinent and beyond;
- Supplements collections of small libraries such as public libraries, school libraries and mobile libraries;
- Supplements national libraries in the sub-continent which have very large collections, but citizens do not have online access to those collections;
- Supplements the community learning centres, village knowledge centres and community e-centres with substantive e-contents and knowledge resources;
- Helps self-learners, lifelong learners and neo-literates increasing their skills, knowledge and expertise;
- Helps researchers in different disciplines through online literature search and retrieval facility;
- Inculcates reading habits to school children and young adults, who have access to and can interact with digital world in a meaningful way;
- Makes locating the relevant information inside of books far more reliable and much easier;
- Increases visibility and availability of Indian books published by Indian institutions and publishers;

- Encourages other resourceful South Asian institutions to digitize and disseminate their knowledge resources through online portals; and
- Creates significant base of South Asian knowledge resources available, searchable and retrievable online.

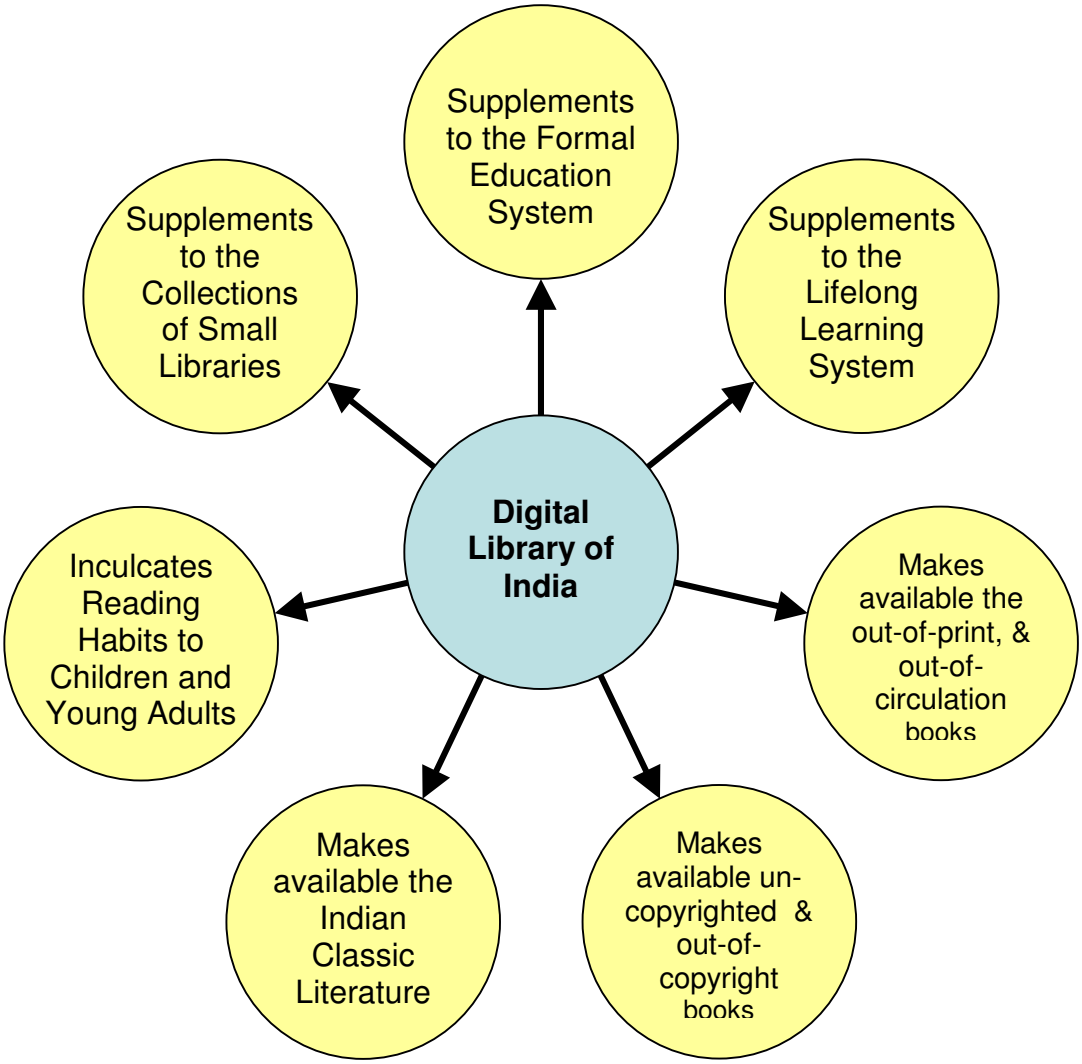


Fig. 2.65: Benefits of Digital Library of India

2.17 DLI – A Test Bed for Indian Language Applications

One of the major objectives of DLI project is to become a repository of public domain Indian language books. Thus, a significant volume of Indian language books were digitized across the scanning centres in India. DLI portals become most resourceful digital library of Indian language books, having several million scanned pages in Indian languages.

As several million scanned pages in Indian languages are available with the DLI scanning centres and RMSCs, DLI became a test-bed for many Indian language applications, spearheaded by DLI research partners – Indian Institute of Science Bangalore (IISc), International Institute of Information Technology Hyderabad (IIIT Hyderabad) and Centre for Development of Advanced Computing (CDAC). These institutions also have collaborated with Technology Development for Indian Languages Programme (TDIL) of Ministry of Communication and Information Technology (MCIT) for development of software applications for Indian language computing and spreading Indian language applications to the cross-section of people.

Some of the Indian language applications developed by DLI research partners and described in different research papers (Shukla, Arora & Gugnani, 2004; Balakrishnan, 2005; Prahallad & Black, 2005; Ganapathiraju, Balakrishnan, Balakrishnan & Reddy, 2005; Prahallad, Prahallad & Ganapathiraju, 2005; Seethalakshmi, R. et al., 2005; Balakrishnan, Reddy, Ganapathiraju & Ambati, 2006), are:

- OCR (Optical Character Recognition) Software for Indian languages, e.g., Kannada OCR.
- TTS (Text-to-Speech) Software for Indian languages.
- Example-based Multilingual Machine Translator for Indian languages, e.g., Saraswati Machine Translation Tool.
- Transliteration Tool for Indian languages, e.g., OM Transliteration Editor.
- Indian Language Search Engine, and
- Indian Language Book Reader.

Figure 2.66 provides a graphical overview of Indian language applications, developed by DLI research partners from DLI datasets.

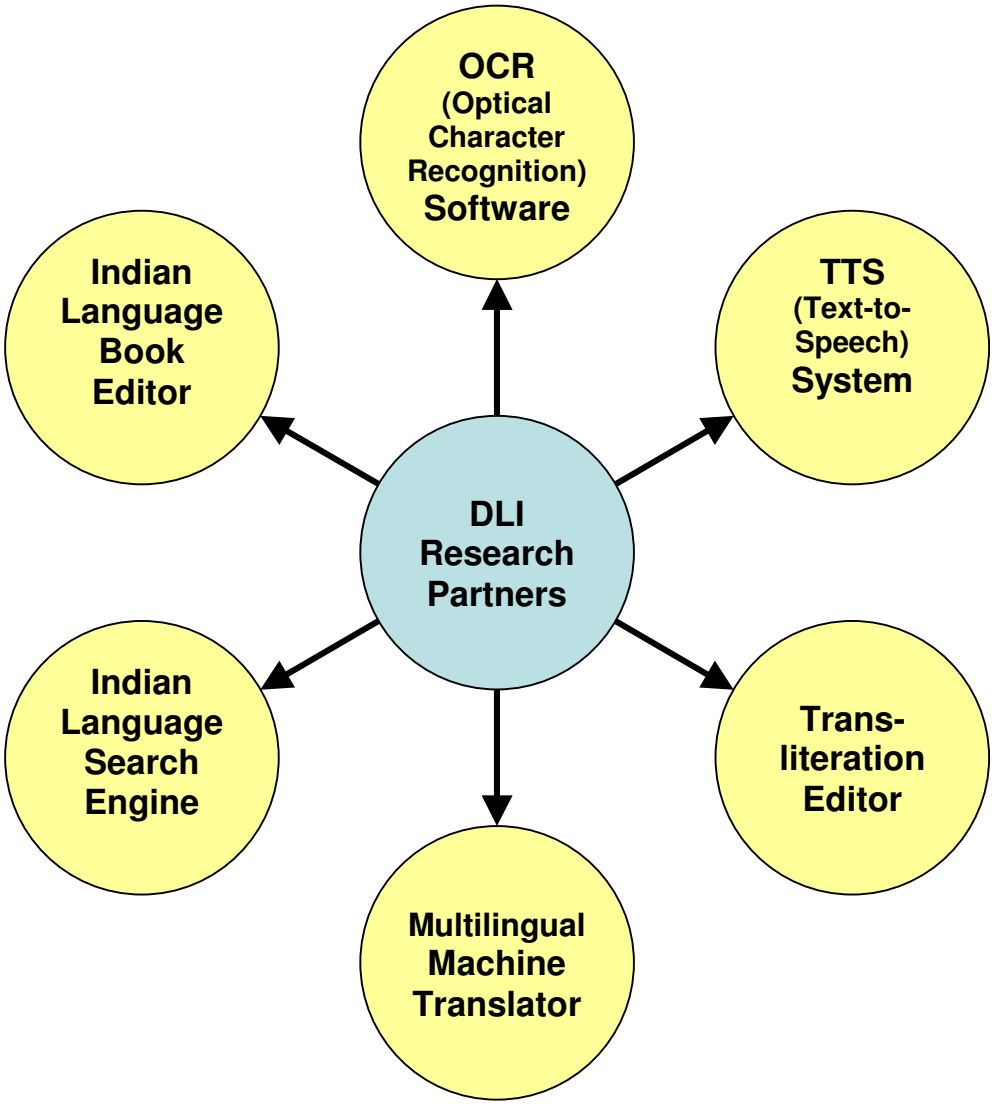


Fig. 2.66: Indian Language Applications Developed by DLI Research Partners

2.18 Some Observations

Metadata is an essential part of any digital library system. In statistical analysis section of this chapter it is found that many metadata elements are missing or wrong metadata information is embedded in metadata databases of DLI systems resulting in non-retrieval of digitized books although available in the DLI system. The metadata quality control should be given more importance and should be performed before Web-hosting of digitized contents.

Three different DLI portals are in existence, having different sets of digitized collections. A single gateway is needed for cross-searching across three portals. This gateway will reduce searching time of the end users.

In DLI portals, page-wise viewing of retrieved books is only possible, resulting in inconvenience to the end users. While in real life digital world end users can consult whole document at any point of time, page-wise retrieval and viewing of digitized books could be avoided in DLI system.

The TIFF file format is also not a standard file format for Web-based system. TIFF images have larger file size and the users having low-bandwidth connectivity face problem of higher page retrieval time. Instead of retrieving many TIFF images in Web browser for viewing a single book, a PDF file could also be provided binding all pages of the book.

First phases of DLI project are now on the verge of completion, where some scanning centres across India have already completed their digitization work and discontinued their digitization services. Technological infrastructure already available with those DLI scanning centres, can be properly utilized to carry out digitization work of other national digitization programmes such as digitization programme of National Mission for Manuscripts.

Capitalizing digitization experience and expertise attained by DLI partner institutions across India, many other mission mode projects can also be undertaken to bridge the digital divide in the country and to harness digital dividend across the stakeholders.

Digital Library of India Initiative follows a decentralized scanning model and centralized structure for project planning, monitoring, evaluation, quality assurance, Web-hosting, technological innovation and technology deployment. DLI also acts as technology incubation centre to develop new technologies and transfer technologies to partner institutions and industries. DLI project also helped in capacity development of partner institutions in general, and national capacity development in particular. Many spin-off technologies are also easily found outside its periphery.

Thus, DLI decentralized structure has proved to be a sustainable model for the benefit of an emerging economy in terms of country-wide capacity, capability and competence development.

2.19 References

Ambati, V. & Reddy, R. (2005). Towards a self healing information system for digital libraries. *Journal of Zhejiang University Science*, 6A(11),1221-1228.

Balakrishnan, N. (2005). Universal Digital Library – future research directions. *Journal of Zhejiang University Science*, 6A(11),1204-1205.

Balakrishnan, N. (2006). *Universal Digital Library: connecting users to digital contents.* Retrieved 15 May 2008 from www.valaconf.org.au/vala2006/papers2006/93_Balakrishnan_Final.pdf

Balakrishnan, N., Reddy, R., Ganapathiraju, M. & Ambati, V. (2006). Digital Library of India: a testbed for Indian language research. *TCDL Bulletin*, 3(1).

Covey, D.T. (2005). *Global cooperation for global access: the Million Book Project.* Retrieved 15 May 2008 from www.library.cmu.edu/People/troll/TrollCoveyBelgiumFinal2r.ppt

Ganapathiraju, M., Balakrishnan M., Balakrishnan N. & Reddy R. (2005). Om: One tool for many (Indian) languages. *Journal of Zhejiang University Science*, 6A(11),1348--1353.

Michalek, G.V., ed. (2003). *Million Book Universal Library Project: manual for metadata capture, digitization, and OCR*. Carnegie: Carnegie Mellon

University. Retrieved 15 May 2008 from www.library.cmu.edu/Libraries/MillionBookManual.pdf

Prahallad, K. & Black, A. (2005). A text to speech interface for Universal Digital Library. *Journal of Zhejiang University Science*, 6A(11), 1229-1234.

Prahallad, L., Prahallad, K. & Ganapathiraju, M. (2005). A simple approach for building transliteration editors for Indian languages. *Journal of Zhejiang University Science*, 6A(11),1354-1361.

Pratha, L., Ambati, V. & Sankar, P. (2005). *Content development for Digital Library of India.* Retrieved 15 May 2008 from www.ulib.org/conference/genpub/dli.pdf

Seethalakshmi, R. et al. (2005). Optical character recognition for printed Tamil text using Unicode. *Journal of Zhejiang University Science*, 6A(11), 1297-1305.

Shukla, V.N., Arora, K.K. & Gugnani, V. (2004). *Digital library: language centered research, test beds and applications.* Retrieved 15 May 2008 from www.cdacnoida.in/technicalpapers/PaperICDL1.pdf

CHAPTER 3

Digitization Work for Theses and Dissertations

3.1 Introduction

India is an emerging knowledge economy, where an array of specialized research institutions, advanced research centres and universities exist in almost all major subject areas. A country worldwide reputed for its IT industries, India is now hosting corporate R&D centres of some major multinational enterprises owing to its global reputation for academic and research excellence. The accreditation and granting agencies for Indian institutions envisaged this phenomenal growth in knowledge-based industries and service industries in India. Thus, these agencies created a structured higher education and research environment in the country in phases at par with global standard. Over the last few decades the Government strengthened ICT and research infrastructure in reputed Indian institutions across the country, where access to knowledge resources and laboratory facilities is made adequate for the students, researchers and faculty members. The government research agencies introduced research fellowship schemes for pursuing research degrees to attract meritorious and talented students into research and teaching professions and for reversing the brain-drain in developed countries. The establishment of open access digital archives of scholarly literature, produced in Indian universities and institutions, is also a kind of ICT intervention that is envisaged to portray India's intellectual capabilities and to ascertain quality of research (Ghosh & Das, 2007). Electronic theses and dissertations (ETD) initiatives and ETD policies in India lie with India's international reputation and to streamline into a systematized and focused research system to generate more social benefits and applications.

3.1.1 Research Fellowship Schemes

The Government of India, through its apex bodies in higher education and research, facilitates research fellowships to support students having aptitude for research. The national agencies such as, University Grants Commission (UGC), Council of Scientific and Industrial Research (CSIR), Indian Council of Medical Research (ICMR), Indian Council of Agricultural Research (ICAR), Department of Biotechnology (DBT) and other national agencies conduct national level eligibility tests for the award of junior research fellowships to ensure minimum standards for entrants in the academic research. The UGC conducts National Eligibility Test (NET) in the areas of humanities, social sciences and applied sciences, whereas CSIR conducts test in the areas of basic sciences (University Grants Commission, 2007). The ICMR, ICAR, DBT also conduct NET in their respective subject areas. The qualified meritorious candidates of NET become eligible to enroll their names in doctoral research programmes in different universities and institutions across the country. Some research fellows then get attached with national laboratories and R&D institutions. The research fellowship is usually available for the first five years of doctoral research, of which first two years as a junior research fellow (JRF) and next three years as a senior research fellow (SRF). In the 11^{th} national Five Year Plan period, which has just started in April 2007 and will continue upto March 2012, Government of India focuses to attract more candidates to junior and senior research fellowships, by achieving flexibility in manpower deployment and enhancement of research fellowship amount. The research fellows are allowed to take part in research projects, consultancy projects and teaching assistantship along side their academic research work and discourse. Presently CSIR inducts about 1200 fresh junior research fellows annually that cumulated to about 7600 research fellows at any point in time, including research fellows employed in CSIR-supported research projects (Planning Commission, 2006). Similarly, UGC, ICMR, DBT, Department of Science & Technology (DST), Indian Council for Social Science Research (ICSSR) and other national agencies together support almost similar number of research fellows in Indian universities and R&D institutions. Table 3.1 depicts the

proposed schemes to support research fellowships in Indian universities and institutions for the 11th Five Year Plan period (Planning Commission, 2006). The responsible agencies also ensure social inclusion and equal opportunities to weaker and backward sections of the society in the research fellowship schemes, including women and physically challenged. All these schemes are meant for increasing national scholarship, knowledge generation capacity and to present India as a self-reliant knowledge-based nation. These fellowship schemes facilitate in producing a sizable number of qualitative knowledge resources, in the forms of doctoral theses, patents, research papers and research reports. As these knowledge resources are emanated from public-funded research, Indian open access crusaders and advocacy groups strongly recommend disseminating the fruits of research through open access channels. Advocacy groups have already succeeded to sensitize government and other stakeholders. Hence, we see a number of open access initiatives in the country, in the forms of ETD repositories, institutional repositories, metadata harvesting services and open access scholarly journals.

Table 3.1: Research Fellowship Schemes available during 11th Five Year Plan

Name of the Scheme	Responsible Agency	Amount in Rupees (Crore)	US Dollar (Million)
Research Fellowships in Basic Sciences & Interdisciplinary areas (JRF/SRF)	CSIR	870.00	206.41
GATE Qualified Junior Research Fellowship (JRF-GATE) Scheme	CSIR	50.00	11.86
Trans-disciplinary Fellowship Scheme	CSIR	14.00	3.32
Shyama Prasad Mukherjee Fellowship Scheme	CSIR	16.00	3.80
Research Fellowships in All Major Subjects & Interdisciplinary Areas	UGC	1500.00	355.86
Teaching Assistantship for Doctoral Students (Non Fellowship)	UGC	150.00	35.87

3.1.2 Policy Frameworks in India

3.1.2.1 UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005

In 2005, the University Grants Commission of India (UGC) drafted a national policy framework entitled "UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005" (UGC, 2005). This Regulation proposed two sets of planned actions, such as:

- **Creation of Indian National Theses Database (INTED)**: The proposed Indian National Theses Database (INTED) is an online centralized bibliographic database, where online submission of metadata sets of a PhD thesis is made mandatory when the researcher finally submits his/her PhD thesis to the Research Cell of a university. PhD supervisor verifies and validates the metadata of submitting PhD thesis. The INTED will be freely accessible worldwide and will have simple and advanced search interfaces. Hence, bibliographic control of all recent PhD theses is assured. Although this online database will be accessible free of charge, its CD-ROM version will be made available at a price.
- **Submission of PhD Theses in Electronic Form***:* Researchers will submit his/her PhD thesis in an appropriate electronic format, along side hard copies of his/ her thesis. The UGC Regulations also suggest a specific content structure of electronic thesis, which is derived from the best practices of international ETD portals. The submitted electronic copy of the doctoral thesis will be stored in the respective University's ETD repository or institutional repository. ETD repositories or institutional repositories should be OAI-PMH compliant, so that metadata harvesting services can collect metadata of each submitted/ awarded doctoral thesis. Administrators of the ETD repository will also check the quality of metadata information. Universities will set up committees to formulate copyright and IPR policies, access policy, plagiarism monitoring policy, and other norms related to ETD repository. Access to ETD database can be allowed to any of the following: worldwide open access,

campus-only access, temporary restricted access, and mixed access (partially open). An e-thesis may contain textual data along with images, audio objects, video objects, animation objects, spatial objects and other kind of multimedia objects. The e-thesis can be navigated non-linearly. Thus, the hard copy of a thesis has less number of functionalities than electronic copy of a thesis. The proposed ETD repositories in universities will be developed using latest versions of IR software and will adopt latest versions of metadata schema to handle the advanced features of ETD.

Implications of UGC Regulations

Proposed set of regulations is still under consideration and stakeholders' consultation or might have been shelved due to unknown reasons. This set of regulations formulates a roadmap in achieving wide dissemination of results of doctoral research conducted in Indian universities and also offers the way of bibliographic control of theses and dissertations submitted for research degrees. So far, a few UGC-supported universities have established open access repositories for scholarly literature produced within those universities. Initially, universities were reluctant to change their status quo. Now, the contents of scholarly literature including the PhD theses will be a matter of critical review by the national and international peers, if the universities establish open access repositories. Gradually, universities will be motivated to produce a large number of qualitative doctoral theses and will maintain certain international standards. The UGC also supports development of infrastructure in Indian universities through various planned schemes. The proposed national education grid will also enrich modern ICT infrastructure in Indian universities. The Indian universities then will have necessary infrastructure to host a number of Web-based information services. Hosting an ETD repository and providing online interface to INTED will not be a problem for most of the universities. Now, Indian universities should take proactive role in implementation of these set of regulations at the earliest, which will ensure qualitative research and make the results of doctoral research widely available. The UGC should also waive all bottlenecks to

implement INTED and decentralized ETD repositories across the country. This way the National Knowledge Commission's recommendation on peer-reviewed research papers resulting from publicly funded research would be validated by making them available through open access channels (NKC, 2006).

3.1.2.2 INDEST-AICTE Consortium – An Open Access Advocacy Group

The Indian National Digital Library in Engineering Sciences and Technology (INDEST) Consortium is an innovative initiative supported by the Ministry of Human Resource Development (MHRD) and All India Council for Technical Education (AICTE). Indian Institute of Technology Delhi (IITD) is the coordinating agency for this consortium. The INDEST-AICTE Consortium tries to address the problems of information poverty in the country and more particularly in its member technical institutions through highly discounted rates of subscription and better terms of agreement with the publishers of electronic resources such as full-text e-journals and scholarly databases. The Consortium maintains three tiers of membership where total membership strength is 741 as on 27 July 2008. Categories of membership are shown below:

- Core Members (supported by MHRD) (37)
  - Indian Institute of Science (IISc)
  - Indian Institutes of Technology (IITs) (7)
  - Indian Institutes of Management (IIMs) (6)
  - National Institutes of Technology (NITs) (17)
  - Indian School of Mines
  - National Institute of Training and Industrial Engineering
  - North Eastern Regional Institute of Science & Technology
  - SHS Longowal Central Institute of Engineering and Technology
  - ABV Indian Institute of Information Technology and Management
- AICTE-supported Members (government engineering colleges) (60)
- Self-supported Institutions (AICTE-accredited and UGC-affiliated institutions) (183)
- Self-supported Institutions (private institutions recognized by AICTE) (461)

Figure 3.1 depicts the membership pattern of INDEST-AICTE Consortium. This consortium maintains a PPP (Public Private Partnership) model for equitable access to scholarly literature.

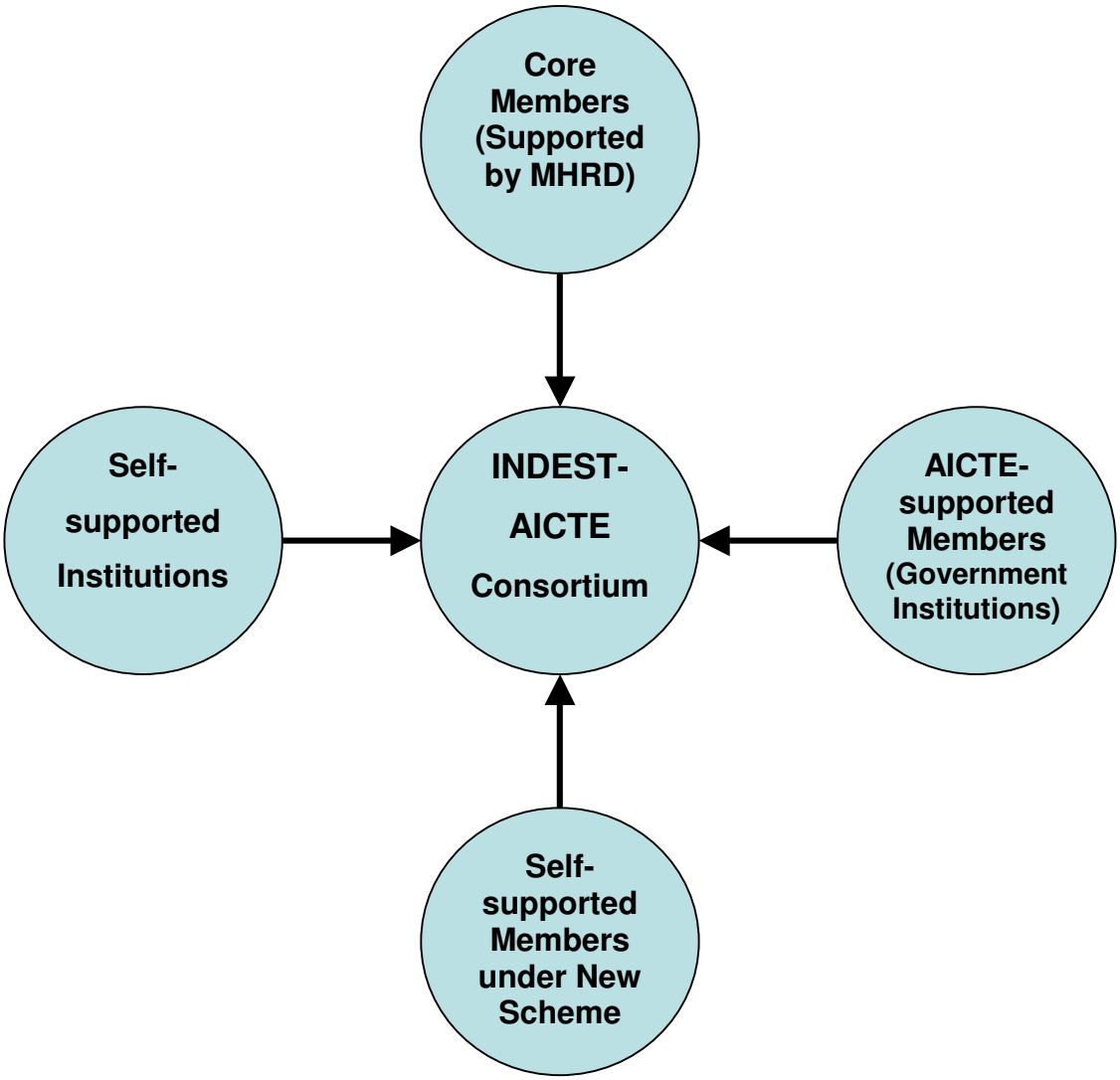


Fig. 3.1: Membership pattern of INDEST-AICTE Consortium

The core member institutions of INDEST Consortium are globally reputed for their academic and research excellence. These institutions have ICT infrastructure and an access to e-resources at par with the best institutions of the world. Some of these institutions have already established open access repositories to make their research literature available globally. INDEST Consortium holds annual meetings to discuss the policies and operational issues. In its 2003 Annual Meeting, INDEST core members proposed to establish an INDEST Digital Theses Archive, where Indian Institutes of Technology and Indian Institute of Science will be major stakeholders. The proposed Digital Theses Archive will have following characteristics (INDEST Consortium, 2003; Venkadesan & Rajashekar, 2003):

- Joint initiative between the member institutions, who constitute INDEST
- Full-text database of the masters and doctoral theses from IITs and IISc
- Distributed national and regional archives
- Facilitate submission of theses and dissertations electronically by the researchers/ students
- Retro-conversion of existing theses to digital form
- Quality of an academic institution will be reflected by the quality of its students' theses and dissertations.

The INDEST will have following roles and tasks as the central coordination agency for national ETD initiative (INDEST Consortium, 2003; Venkadesan & Rajashekar, 2003):

- coordinating and cooperating with member institutions on organizational, technological and educational issues and for developments;

- fund raising, supporting and conceptualizing of decentralized ETD structure, special interest groups and working groups in which representatives from different institutions can participate;

- organizing workshops for participating member institutions covering general and special topics on ETD formats and archiving methods; and

- liaison with other Indian initiatives such as 'Vidyanidhi' and international initiatives such as NDLTD.

Text Box 3.1: Recommendations of INDEST Working Group on 'ETD and EPrint Archives'

ETD and EPrint Archives

1. All IITs and IISc should implement electronic submission of theses and dissertations in addition to their current practice of submissions of theses and dissertations in print;
2. Following the model set-up by the IISc Bangalore, all IITs may set-up e-print archives using OAI-compliant e-print software. While pre-prints and re-prints may be submitted at e-print servers set-up at various IITs / IISc, a central server may be deployed to harvest the metadata from all such e-print archives. The use of OIA-compliant e-print software would facilitate such an environment. It was strongly felt that if the pre-print / re-print submission is followed strictly at all IITs / IISc, the unified collection developed in the process at IITs and IISc (as well as at other such repositories) would offer a strong alternative to existing model of scholarly publishing;
3. All INDEST Consortium members (i.e. IITs and IISc) should follow common protocols and procedures to ensure interoperability of digital collections built-up in the process of e-submissions of theses, dissertations and preprints. The IITs and IISc may adopt OAI-compliant software available in public domain, such as e-prints archive, D-space or the Virginia Tech.
4. The ETD may be planned as distributed archives with individual institutions holding their theses and dissertations on their own servers while the metadata may be harvested either online or offline on a single server;
5. IIT Bombay and IIT Kanpur, the two institutions who have already implemented ETD at their respective institutions, may provide trial access to their ETDs to other IITs / IISc;
6. A core development group may be formed to give the technical specifications for the ETD as well as for preprints and for the development of the archive site. This group will also prepare a template for the ETD and preprint site, which can be used by any of the INDEST members to host their ETD archive. The Group may also finalize standards for the metadata needed for the ETD.

Source: Minutes of the Meeting held on 7th October, 2003 at IIT Delhi
http://paniit.iitd.ac.in/indest/extended/minutes7thoct.pdf

Implications of INDEST Recommendations

Text Box 3.1 shows a detailed account of recommendations of INDEST Working Group on 'ETD and EPrint Archives'. Some core member institutions have already established open access institutional repositories and ETD repositories in their respective institutions. Some core member institutions such as IISc, IITs, have started digitization of their back volumes of thesis and dissertation collections on project basis and making them available through open access or campus-wide institutional/ETD repositories. ETD@IISc is considered as a role model of open access ETD repository in the INDEST family. INDEST Consortium in association with IISc has prepared a number of guidelines on the ETD and organized training workshops for implementations of open access repositories in member institutions. This Consortium also provides consultative supports to other Indian consortiums, research agencies and individual institutions in formulation of open access policies and implementation of ETD/institutional repositories.

3.1.2.3 Doctoral Theses Repository in NASSDOC

The National Social Science Documentation Centre (NASSDOC) is a constituent institution of Indian Council of Social Science Research (ICSSR) and a national repository of doctoral theses in the areas of social sciences. NASSDOC is also an official partner of Vidyanidhi. As a national library of social sciences, NASSDOC aims to build a truly representative collection of doctoral theses and dissertations in all areas of social sciences, including interdisciplinary areas. NASSDOC systematically collects qualitative doctoral theses in the areas of social sciences submitted in Indian universities through an incentive scheme called 'Acquisition of Thesis'. So far NASSDOC has acquired about 5000 doctoral theses in different areas of social sciences in its special collection. For consideration in this acquisition scheme, the researcher sends the details of his/her research work, i.e., synopsis containing scope, methodology, objectives and major findings. A screening committee examines details of the thesis and recommends decision on acquisition. Doctoral theses on India by any Indian or

foreign students approved by the foreign universities are also acquired through this scheme. NASSDOC has planned to digitize and microfilm its theses collection to address space problem and to device a modern retrieval mechanism of social sciences information. The integration of its digitized ETD collection into Vidyanidhi is a right direction to diffuse Indian social science research findings to worldwide audience. But NASSDOC should keep in mind that Vidyanidhi may be a platform to start with, but NASSDOC should establish an alternative dedicated server for any long-term preservation and access.

3.1.2.4 Recommendations of National Knowledge Commission

The National Knowledge Commission of India (NKC), constituted on 13th June 2005, is a high-level advisory body to the Prime Minister of India, with a mandate to guide policy and direct reforms. NKC's overarching aim is to transform India into a vibrant knowledge-based society. In its first annual report entitled 'Report to the Nation 2006', NKC published its first set of recommendations to the PM of India (NKC, 2006). Subsequently, NKC published its second annual report entitled 'Report to the Nation 2007' having another set of recommendations for open educational resources and other knowledge-based areas (NKC, 2007). If implemented, these recommendations will have far-reaching implications in the knowledge creation and dissemination cycle. NKC strongly advocates open access to public-funded research literature (NKC, 2006). The theses and dissertations produced by state-sponsored researchers would then be made accessible through open access channels such as ETD repositories or institutional repositories. NKC also recommends for the expansion of Indian higher education system through establishment of new universities, including 50 national universities, totaling about 1500 universities across the country, which is about 3 times addition to existing system. This expansion would enable India to attain a gross enrolment ratio in higher education at least 15 per cent by 2015. This kind of phenomenal expansion requires a huge pool of qualified teaching and research staff, equipped with higher research degrees, to maintain a minimum standard of teaching and research. Similarly, the present university

system should also take proper measures to attract meritorious research students for future requirements and disseminate their research papers, theses and dissertations by establishing ETD/institutional repositories.

3.2 Open Access to Full-text Theses and Dissertations in India

India is spearheading open access movement in the developing countries since last decade by establishing a number of open access repositories, embracing free and open source software (FOSS). Indian information professionals experimented with the open source software such as Greenstone, DSpace and EPrints, for establishing institutional repository (IR) in a local library. If an IR is successfully implemented in the local library setup, it then scales up to the institution-wide application through campus-wide network or intranet. Similarly it turns open to the wider audience with the implementation of open access institutional repository, when the authority of institution is convinced. With the availability of the dedicated information infrastructure combined with 24X7 broad band connectivity and a national educational grid, some national institutions and universities implemented institutional repositories for wide dissemination of scholarly literature emanated from the respective institutions. Some institutional repositories in India are specially established to diffuse intellectual outputs of the country in the form of electronic theses. Vidyanidhi and ETD@IISc are examples of such kind. Other open access repositories provide all kinds of scholarly materials such as research papers, conference papers, presentations, photographs, along with e-theses. The OpenMED and Librarians' Digital Library are examples of this kind. Figure 3.2 illustrates types of ETD repositories exist in India. Open access repositories depicted in this Figure are either operated as a national level subject repository or a national level ETD repository or an institutional repository or an institutional ETD repository. National-level subject repositories and institutional repositories have mixed collections of scholarly literature, including an ETD collection. Whereas, national level ETD repository or institutional ETD repository only contains ETD collection excluding all other forms of scholarly publications.

Table 3.2 provides an indicative list of open access repositories in India having ETD collections. These ETD repositories provide worldwide access to results of Indian research. Thus, these repositories are listed in one or many international directories/harvesters of open access repositories such as OpenDOAR (Directory of Open Access Repositories), OAIster, ROAR (Registry of Open Access Repositories), Google Scholar and CASSIR (Cross Archive Search Service for Indian Repositories).

Table 3.3 provides a quantitative list of collections in different open access repositories, indicating the number of ETD available in each repository vis-à-vis other scholarly items. This Table also provides statistical information as on two dates with 14 months gap. This time gap also helps us to know growth rate of ETD collections in the listed repositories. This Table indicates that Vidyanidhi stores maximum number of full-text e-theses followed by DSpace@TIET, ETD@IISc and DSpace@NCL as on 15 May 2007. This Table further shows that Vidyanidhi stores maximum number of full-text e-theses followed by DSpace@TIET, ETD@IISc and RRI Digital Repository as on 21 July 2008. Unfortunately, institutional repository of National Chemical Laboratory (DSpace@NCL), which stood 4th in 2007, went offline in 2008 due to some changes in institutional policy. Table 3.3 identifies that top three open access digital archives are showing a steady rate of growth, i.e., 31.97% , 60.23% and 29.69% respectively in terms of total number of full-text e-theses. This Table also indicates that Indian repositories experiencing more than 100 percent growth are RRI Digital Repository, OpenMED@NIC, and DSpace@NITR.

Figure 3.3 graphically depicts ETD collections in Indian open access repositories, based on Table 3.3. This Figure also compares ETD collections vis-à-vis other items available in Indian open access archives. Amongst all open access archives in India, Vidyanidhi has maximum number of full-text items followed by RRI Digital Repository, and Open Access Repository of IIAP.

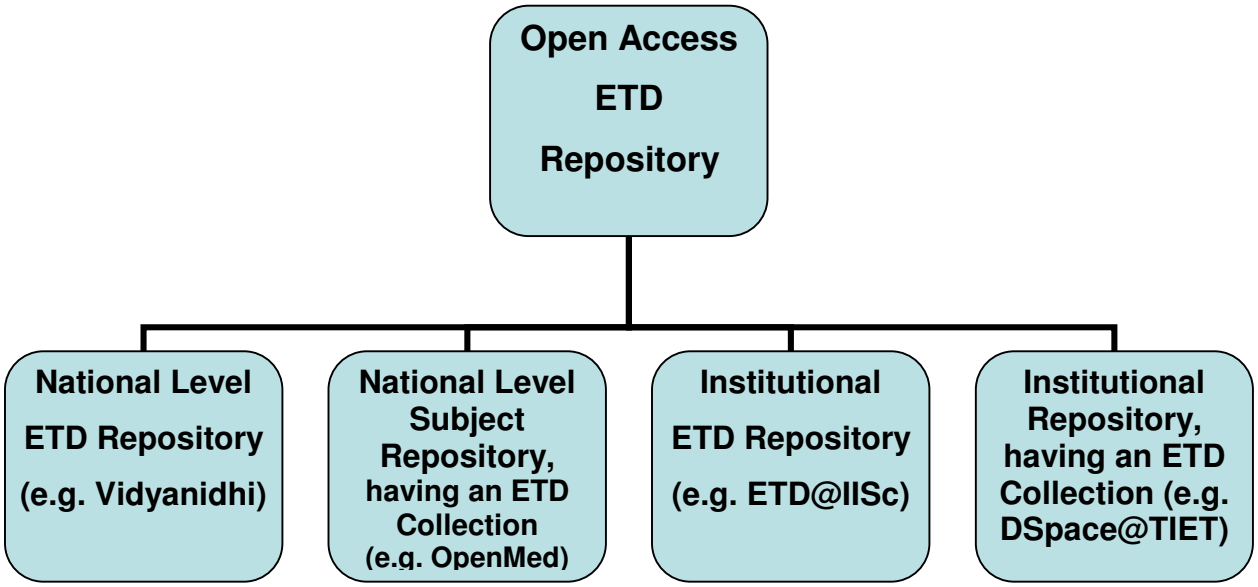


Fig. 3.2: Models of Indian Open Access ETD Repositories

Table 3.2: Indian Open Access Repositories having ETD Collections

Name or the Repository	Name of Institution	URL	Software Used
Vidyanidhi	University of Mysore	http://dspace.vidyanidhi.org.in:8080/dspace/	DSpace
Dspace@TIET	Thapar University	http://dspace.tiet.ac.in:8080/dspace/	DSpace
ETD@IISc	Indian Institute of Science	http://etd.ncsi.iisc.ernet.in/	DSpace
DSpace@NCL	National Chemical Laboratory	http://dspace.ncl.res.in/dspace/	DSpace
Open Access Repository of IIAP	Indian Institute of Astrophysics	http://prints.iiap.res.in/	DSpace
Eprints & ETD@IIT Delhi	Indian Institute of Technology Delhi	http://eprint.iitd.ac.in/dspace/	DSpace
DSpace@NITR	National Institute of Technology Rourkela	http://dspace.nitrkl.ac.in/dspace/	DSpace
OpenMED@NIC	National Informatics Centre	http://openmed.nic.in	EPrints
Librarians' Digital Library	Documentation Research and Training Centre	https://drtc.isibang.ac.in/	DSpace
IIMK's Scholarship Repository	Indian Institute of Management Kozhikode	http://dspace.iimk.ac.in/	DSpace
RRI Digital Repository	Raman Research Institute	http://dspace.rri.res.in/dspace/	DSpace
DSpace at CUSAT	Cochin University of Science & Technology	http://dspace.cusat.ac.in/	DSpace

Table 3.3: Distribution of ETD Collections in Indian Open Access Repositories

Name or the Repository	ETD as on 21/07/08	ETD as on 15/05/07	Other Items as on 21/07/08	Other Items as on 15/05/07	Total Collection as on 21/07/08	ETD Growth Rate (%)
Vidyanidhi	5478	4151	0	0	5478	31.97
Dspace@TIET	415	259	59	33	474	60.23
ETD@IISc	297	229	0	0	297	29.69
DSpace@NCL	0	208	0	149	0	-100.00
Open Access Repository of IIAP	85	83	2590	1389	2675	2.41
Eprints & ETD@IIT Delhi	30	30	2141	2106	2171	0.00
DSpace@NITR	36	13	635	388	671	176.92
OpenMED@NIC	51	8	2095	1491	2146	537.50
Librarians' Digital Library	5	3	376	364	381	66.67
IIMK's Scholarship Repository	2	2	395	256	397	0.00
RRI Digital Repository	95	1	3383	2191	3478	9400.00
DSpace at CUSAT	39	-	1383	-	1422	-

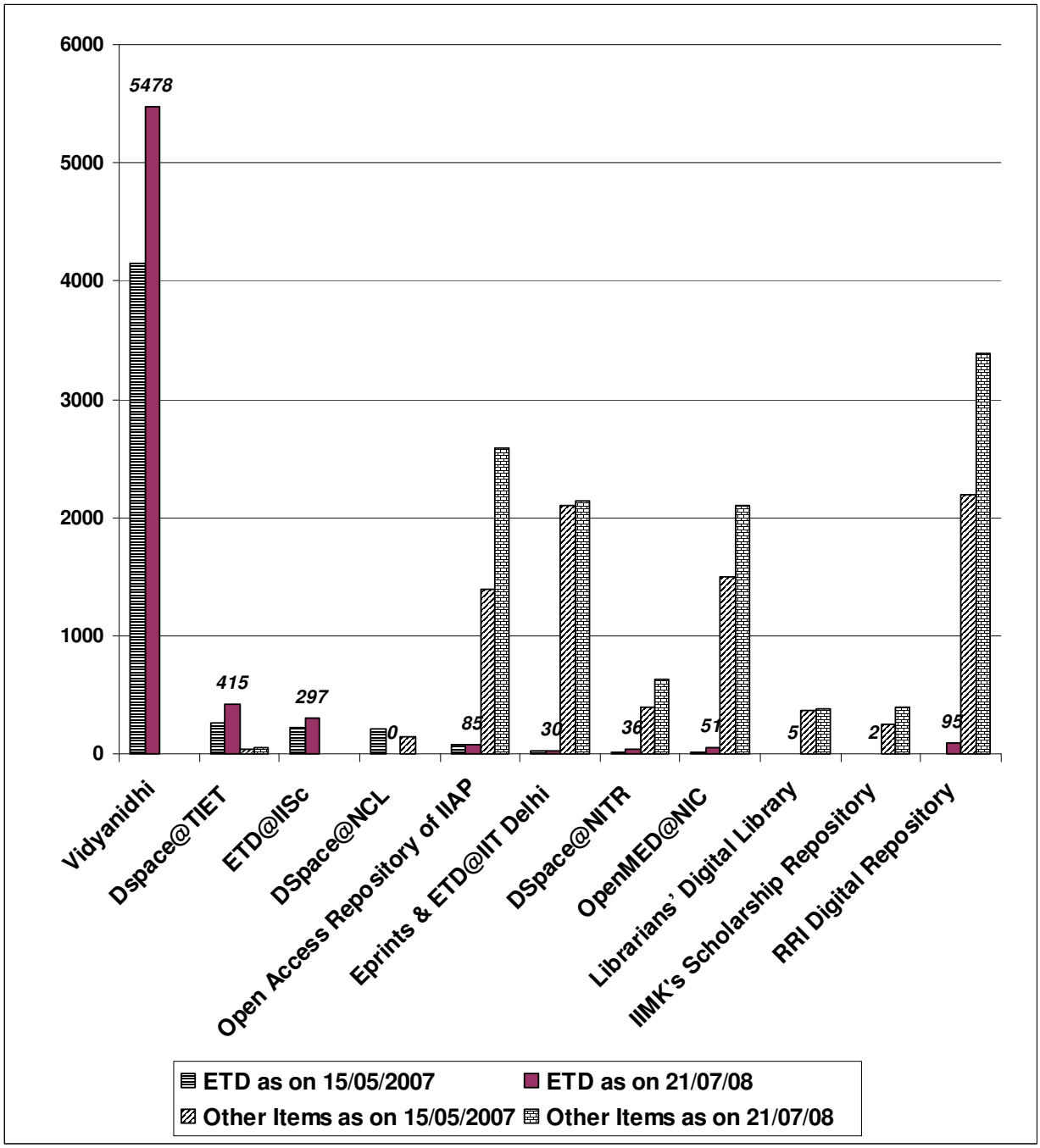


Fig. 3.3: ETD Collections in Indian Open Access Repositories

3.3 Digitization of Theses and Dissertations

An ETD repository contains both digitized theses and born digital theses. Physical collections of theses and dissertations in university libraries are being digitized using appropriate digitization technologies such as book scanners, image processing software, OCR (optical character recognition) software, metadata entry forms, and XML style sheets. The process of digitization of theses is similar to book digitization process. Scanned theses and dissertations, written in English language, can be converted into texts through the OCR process and textual files are stored in searchable-PDF file format. On the other hand, scanned theses and dissertations, written in Indian languages, are usually stored in non-searchable PDF file format, retaining images of scanned pages. Usually, digital files of theses and dissertations are created chapter-wise. That means a record of a thesis or dissertation in the ETD repository links to several files for different chapters as shown in Figure 3.5.

Indian universities now-a-days accept electronic copies of theses and dissertations along with print copies. A research scholar submits electronic version of his thesis or dissertation to degree granting university and/or to institution, where he carried out his research work. Figure 3.4 depicts an ETD repository that contains born-digital theses and dissertations which are contemporary in nature. ETD repository can also contain digitized theses and dissertations if the host university converts physical collections into electronic objects through a digitization project. Some Indian ETD initiatives have undertaken time-bound digitization project for converting physical collections of theses and dissertations into electronic format. Vidyanidhi digital library, Indian Institute of Science Bangalore (IISc), Indian Institute of Technology Delhi (IIT Delhi) are the most suitable examples in this effort, where significant volumes of theses and dissertations have been converted into digital form for providing worldwide access to Indian theses literature.

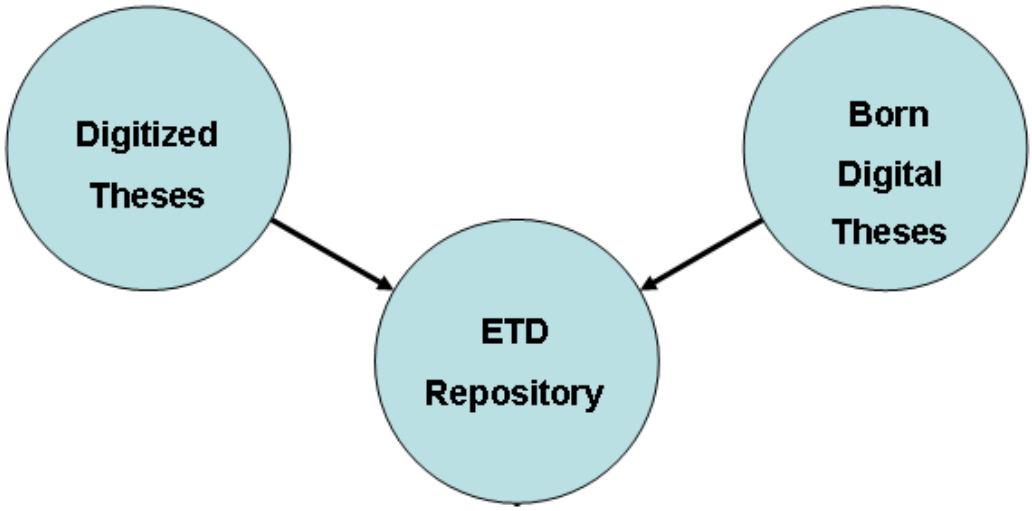


Fig. 3.4: Making of an ETD Repository

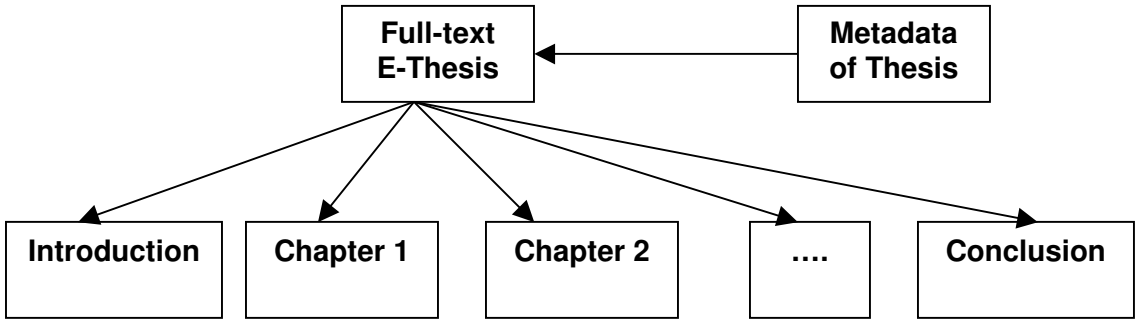


Fig. 3.5: Tree Structure of Contents of Full-text e-Thesis

In the following sections of this chapter, three most significant ETD initiatives, namely, Vidyanidhi Digital Library, ETD@IISc and OpenMED@NIC, are described along with an analysis of their impact on the research communication, international visibility and availability of theses literature. These initiatives are replicable and scalable. Other higher educational institutions in the country can either collaborate with these two national ETD initiatives, or establish an

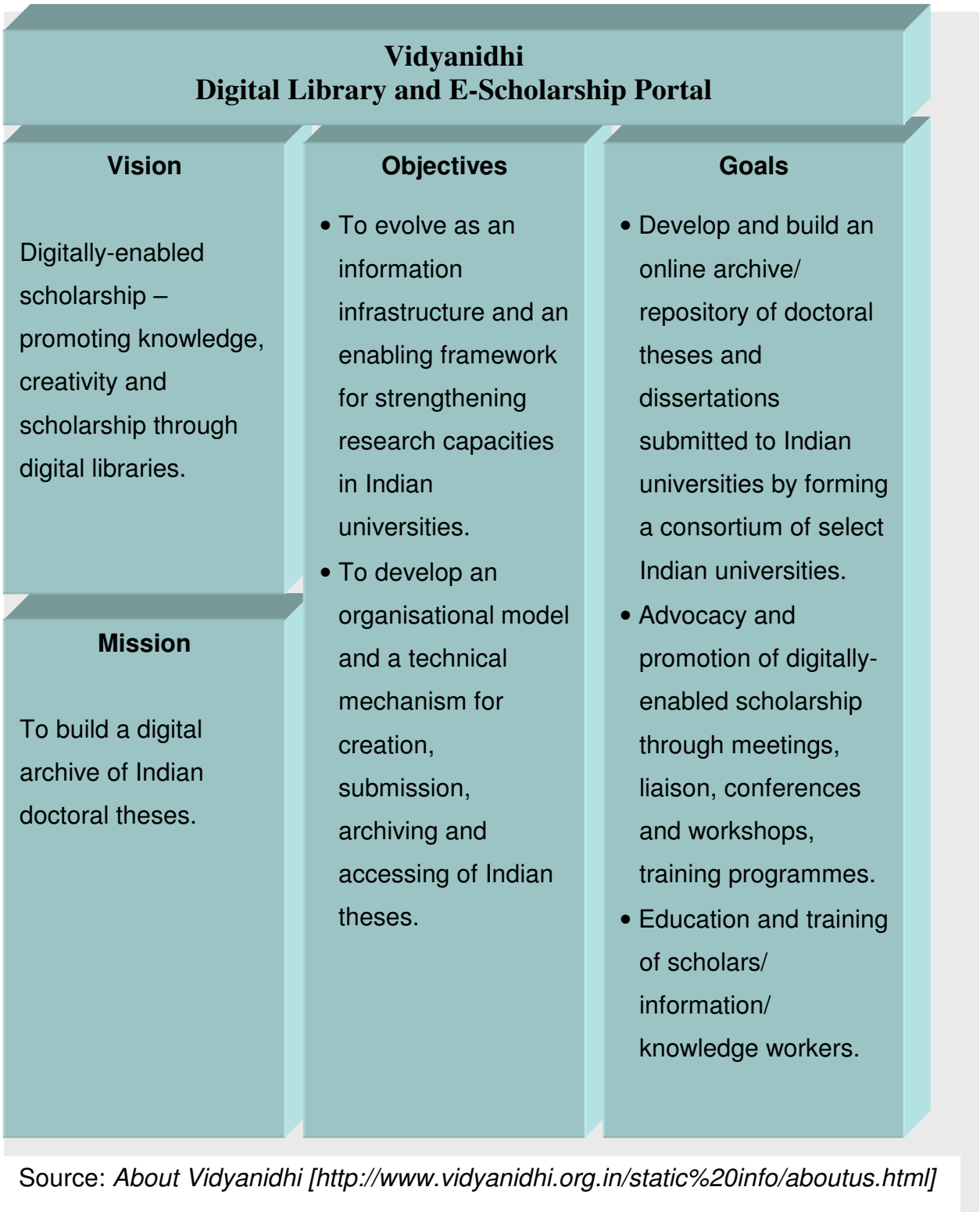
institutional ETD repositories similar to ETD@IISc for worldwide dissemination of their research results.

3.4 Vidyanidhi Digital Library Initiative [www.vidyanidhi.org.in]

Vidyanidhi Digital Library is one of the internationally known Indian digital libraries, initiated in 2000. This is a national level digitization initiative for theses and dissertations, which is largest in South Asia. Spreading all over the country, establishing a network of five partner institutions and few hundred universities across India, Vidyanidhi maintains an open access repository of Indian scholarship. In this project, theses and dissertations available in physical form are converted into digital form through the process of digitization. Vidyanidhi is experimenting with various types of theses materials and has established the procedure for treatment. The materials digitized in this project are being Web-enabled for wider access. The technology is also transferred to various research degree awarding agencies for information handling and dissemination in a distributed mode.

Vidyanidhi initiative has an overarching mission to build a digital archive of Indian doctoral theses. It envisages digitally-enabled scholarship – promoting knowledge, creativity and scholarship through digital libraries. Vidyanidhi statements on mission, vision, objectives and goals are depicted in Figure 3.6.

Vidyanidhi aims at: (i) enhancing international visibility of Indian university research, (ii) enlarging acess to doctoral research outputs, (iii) expanding audience of research literature, (iv) empowering students to have more citations, (v) enabling universities to produce quality research, (vi) extending dissemination of doctoral research results, (vii) enduring archiving of submitted theses and dissertations, and (viii) enriching expressiveness in doctoral research outputs. These aims of Vidyanidhi initiative are graphically shown in Figure 3.7.



Source: *About Vidyanidhi [http://www.vidyanidhi.org.in/static%20info/aboutus.html]*

Fig. 3.6: Vision, Mission, Objective and Goal Statements of Vidyanidhi Digital Library Project

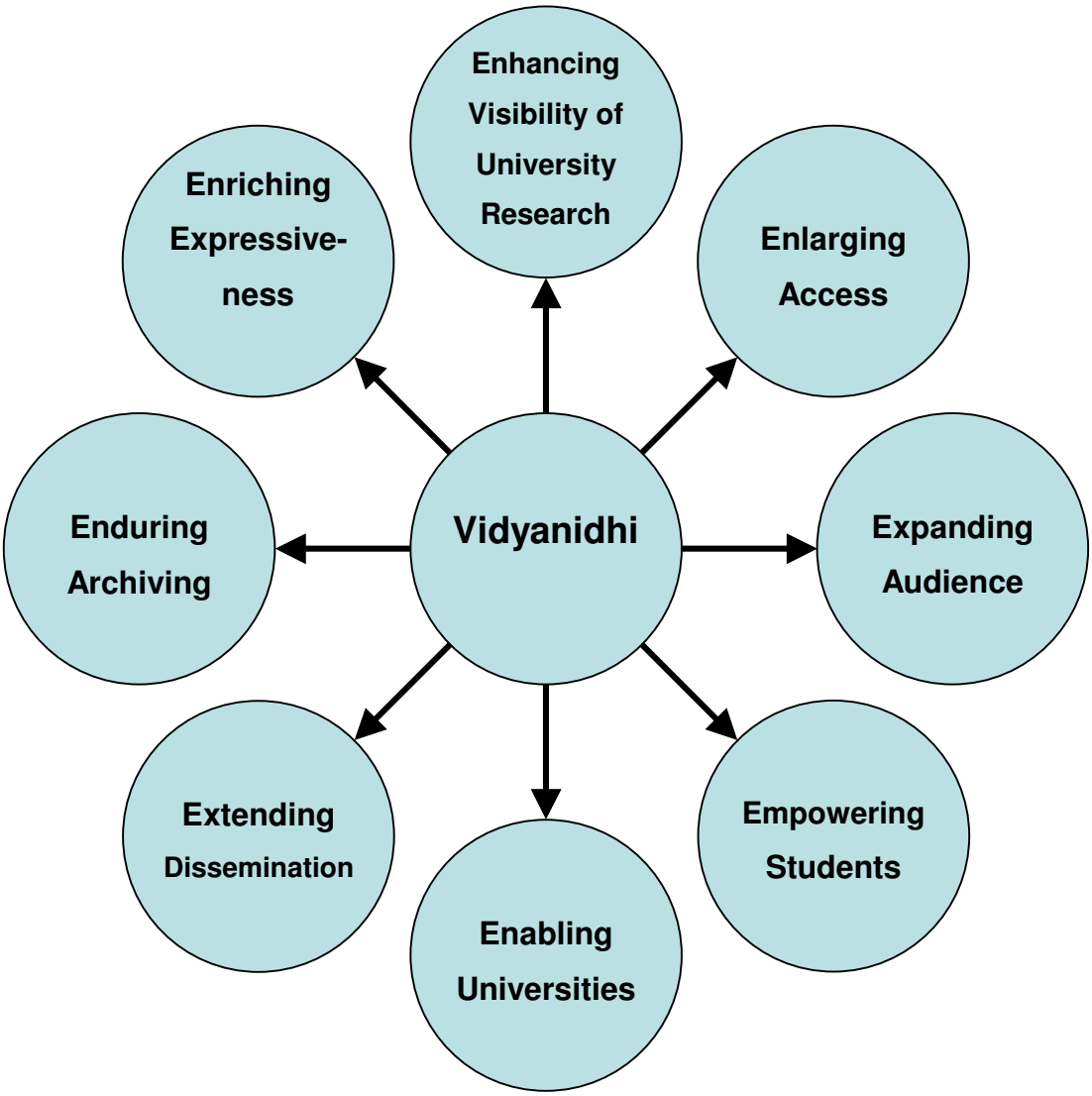


Fig. 3.7: Aims of Vidyanidhi

3.4.1 Partnership Pattern in Vidyanidhi Digital Library Project

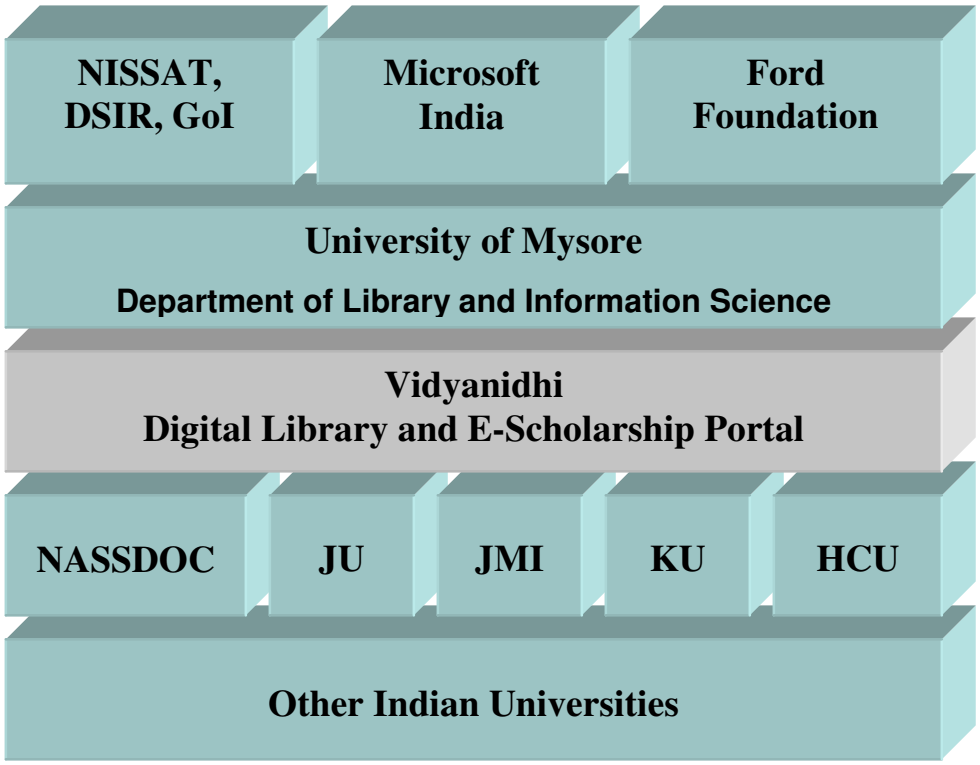
Vidyanidhi Digital Library is a national level digital library initiative for ETD resources, strategically planned, designed and implemented by Department of Library and Information Science at University of Mysore. This project was initially supported by National Information System for Science and Technology (NISSAT), Department of Scientific and Industrial Research of India (DSIR), Government of India, and started in 2000 as a pilot project to develop prototype digital library for doctoral theses. Ford Foundation and Microsoft India are other two funding partners in this project that helped in capacity and infrastructure development for the project. Vidyanidhi is affiliated to Networked Digital Library of Theses and Dissertations (NDLTD) – a global body for promotion and networking of ETD initiatives across the world. As doctoral research conducted in Indian universities is driven by policy frameworks of University Grants Commission (UGC), UGC becomes a natural strategic partner of Vidyanidhi project. In fact, Vidyanidhi played an influential role when UGC drafted a policy framework titled 'UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005', which is still under consideration.

Vidyanidhi establishes two tiers of partnership with Indian institutions – (i) partner institutions, and (ii) participating universities. Partner institutions of Vidyanidhi have more strategic and organic links with Vidyanidhi project. Each partner institution has signed a Memorandum of Understanding (MoU) for establishing decentralized consortia-based content creation and sharing model. Jadavpur University, University of Hyderabad, University of Kashmir, Jamia Millia Islamia and National Social Science Documentation Centre (NASSDOC) are the partner institutions of Vidyanidhi project, as officially recognized in its portal. Other Indian universities also became contributing members in this initiative, for voluntary contribution of their ETD resources into the Vidyanidhi ETD portal.

Figure 3.8 provides a schematic structure of Vidyanidhi project partners. In the top layer, project funding agencies are shown, followed by names of implementing agency, partner institutions and participating universities.

Figure 3.9 identifies partnership pattern in Vidyanidhi project. This Figure also indicates that Vidyanidhi has global partnership with Networked Digital Library of Theses and Dissertations (NDLTD) and strategic partnership with India's University Grants Commission (UGC). While all members of Vidyanidhi community are contributing partners to the initiative, NDLTD stands a clear distinction. Vidyanidhi is connected to this global ETD coordination body as a member.

Table 3.4 provides a comprehensive list of Vidyanidhi project partners, including the institutions shown in Figures 3.8 and 3.9.



NISSAT = National Information System for Science and Technology
DSIR = Department of Scientific and Industrial Research
GoI = Government of India
Partners:
NASSDOC = National Social Science Documentation Centre
JU = Jadavpur University
JMI = Jamia Millia Islamia
KU = University of Kashmir
HCU = University of Hyderabad

Fig. 3.8: Schematic Structure of Vidyanidhi

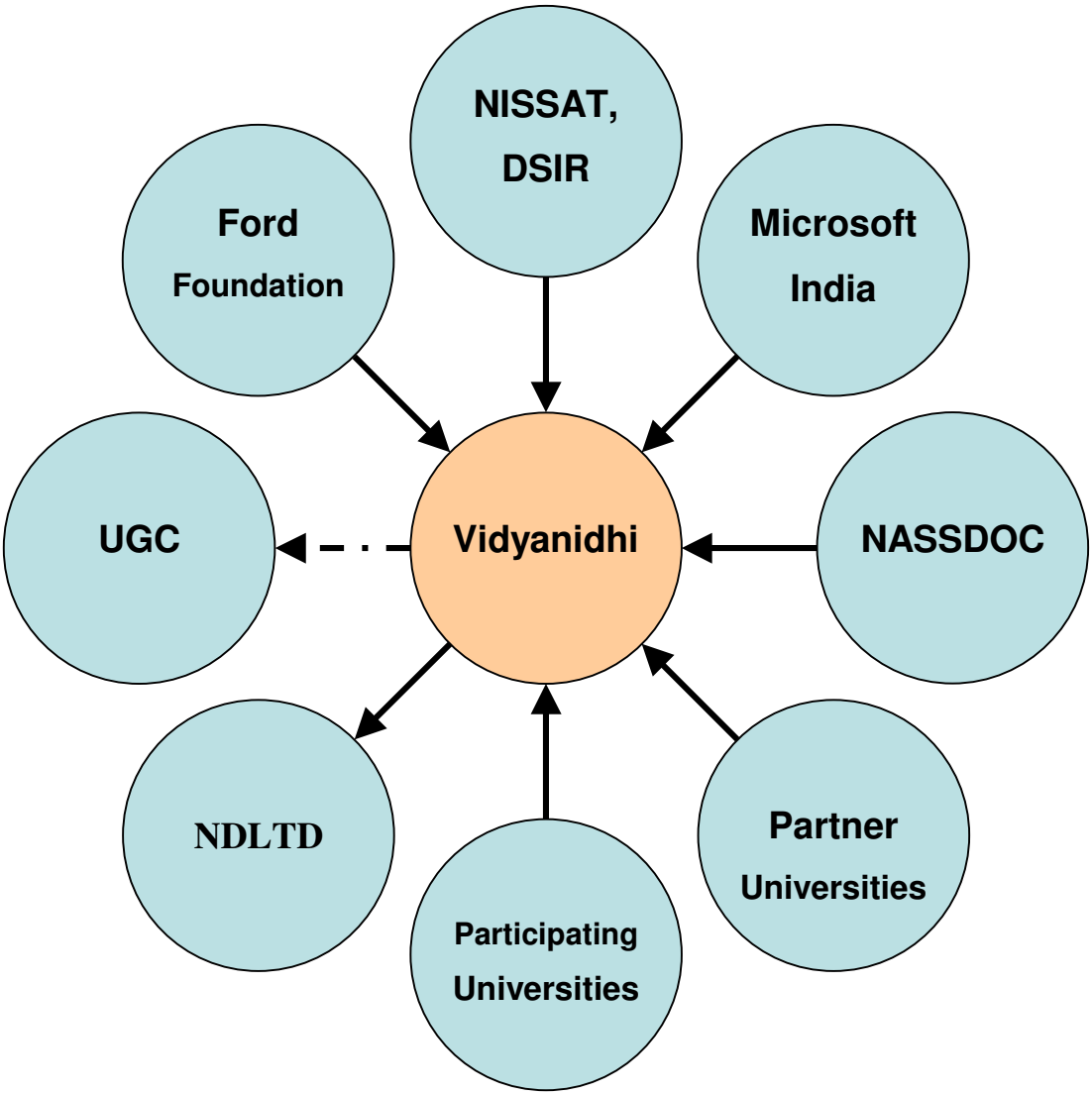


Fig. 3.9: Partnership Pattern in Vidyanidhi Project

Table 3.4: Comprehensive List of Vidyanidhi Project Partners

Name of Vidyanidhi Partner	State	Role	Type of Organization
Department of Library and Information Science, University of Mysore [www.uni-mysore.ac.in/]	Karnataka	Project Directorate	Academic
National Information System for Science and Technology (NISSAT), Department of Scientific and Industrial Research (DSIR) [www.dsir.gov.in/]	Delhi	Funding	Government
Ford Foundation [www.fordfoundation.org/]	Delhi	Funding	Donor Agency
Microsoft India [www.microsoft.com/india/]	Karnataka	Funding	Corporate
University Grants Commission (UGC), India [www.ugc.ac.in/]	Delhi	Strategic Support	Government
Networked Digital Library of Theses and Dissertations (NDLTD) [www.ndltd.org/]	United States	Strategic Support	Global ETD Initiative
National Social Science Documentation Centre (NASSDOC), Indian Council of Social Science Research (ICSSR), New Delhi [www.icssr.org/]	Delhi	Partner Institution	Government
Jadavpur University, Kolkata [www.jadavpur.edu/]	West Bengal	Partner University	Academic
Jamia Millia Islamia, New Delhi [www.jmi.nic.in/]	Delhi	Partner University	Academic
University of Kashmir, Srinagar [www.kashmiruniversity.net/]	Jammu & Kashmir	Partner University	Academic
University of Hyderabad [www.uohyd.ernet.in/]	Andhra Pradesh	Partner University	Academic
Other Universities	All over India	Content Provider	Academic

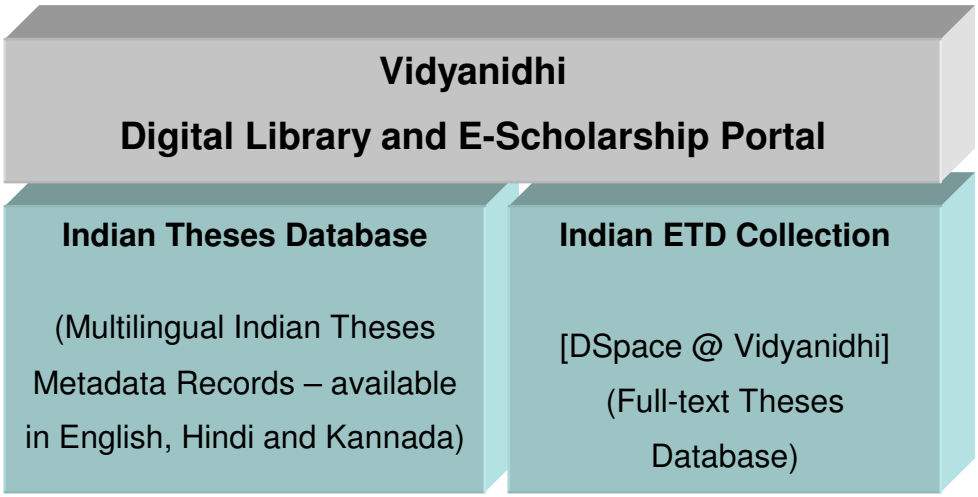


Fig. 3.10: Theses Database and ETD Collection in Vidyanidhi Project

3.4.2 Online Databases of Vidyanidhi

Vidyanidhi is a national level repository, covering a number of universities and deemed universities. Vidyanidhi maintains mainly two databases, 'Indian Theses Database' – an online bibliographic database, 'DSpace@Vidyanidhi' – a full-text repository using DSpace software having an 'Indian ETD Collection', as shown in Figure 3.10. Its bibliographic database has interfaces in English, Hindi and Kannada languages with simple search, advanced search, field-specific search, subject search, supervisor search, and author search facilities. Vidyanidhi also maintains an experts' database, where details of university faculties can be searched by the name of the guide or research area.

Vidyanidhi has an array of partner institutions and participating universities which are instrumental for collection development in its full-text database. Table 3.5 shows the distribution pattern of full-text theses collection by the partner universities. This Table also indicates that University of Mysore is the top contributor, followed by University of Hyderabad and Jamia Millia Islamia. Although Jadavpur University is a partner institution in this effort, it has contributed little as compared to other participating universities.

Vidyanidhi provides an Online Metadata Submission Form to its participating and partner universities, with following metadata elements: Thesis Title, Author's

Name, Guide's Name, Language, Degree Grantor, Department Name, and Submitted Year. All the fields in this online form are mandatory. Table 3.6 indicates metadata elements to be entered through the online metadata submission form.

Indian Theses Database has many search interfaces, including simple and advanced search ones. The search result from Indian Thesis Database displays full details of each thesis with following metadata elements: Title, Creator (Author), Subject, Contributor (Supervisor), Language, Degree Grantor and Year. Table 3.7 shows metadata elements used in Indian Theses Database such as searchable metadata elements and displayed metadata elements. This Table also indicates many search interfaces available to explore this online database.

Table 3.8 illustrates advanced search features of Indian Theses Database along with some suitable examples. This Table also describes how Boolean operators are used in search queries. For example, search query ‘with all of the words’ AND operator is used, search query ‘with exact phrase’ AND operator is used, search query ‘with at least one word’ OR operator is used, and lastly search query ‘without the words’ NOT operator is used.

However, Indian ETD Collection at DSpace@Vidyanidhi uses different sets of metadata elements as inbuilt in DSpace software such as Title, Authors, Subject, Keywords, Date of Submission, Abstract, URI (Uniform Resource Identifier), and Name of Collection. Here author refers to both researcher (who carried out a thesis work) as well as supervisors. The record display page of individual thesis provides links to full-text objects, along with metadata information. Vidyanidhi also maintains an XML sheet for each submitted thesis. Table 3.9 provides a comprehensive list of metadata elements used in DSpace@Vidyanidhi that includes Dublin Core metadata recorded in XML file, browsable metadata, searchable metadata and displayed metadata.

Table 3.5: Distribution of Full-text Theses by Partner Universities of Vidyanidhi

Partner Institution	No. of Theses as on 21/07/2008	No. of Theses as on 15/05/2007
University of Mysore	1805	1657
University of Hyderabad	396	394
Jamia Millia Islamia	271	24
Jadavpur University	4	3
University of Kashmir	3	3
Others	2999	2070
Total	5478	4151

Source: *Our Partners www.vidyanidhi.org.in/asp/ourpartners/ourpartners.aspx*

Table 3.6: Metadata Elements in Online Metadata Submission Form

Name of Metadata Element	Whether Mandatory Field
Thesis Title	Yes
Author's Name	Yes
Guide's Name	Yes
Language	Yes
Degree Grantor	Yes
Department Name	Yes
Submitted Year	Yes

Table 3.7: Search and Display of Metadata Elements in Indian Theses Database

Search Option	Searchable Metadata	Displayed Metadata
➢ Simple Search ➢ Advanced Search ➢ Field Specific Search ➢ Author Search ➢ Supervisor Search ➢ Subject Search ➢ University Search	➢ Thesis Title ➢ Creator ➢ Contributor ➢ Subject ➢ Language ➢ University Name ➢ Year ➢ Keyword	➢ Title ➢ Creator ➢ Subject ➢ Contributor ➢ Language ➢ Degree Grantor ➢ Year

Table 3.8: Advanced Search Option in Indian Theses Database

Search Option	Boolean Operator	Example
Find Results		
➢ With all of the words	➢ AND	➢ **Economic** Development in **India**
➢ With exact phrase	➢ AND	➢ **"Institutional Repositories"**
➢ With at least one word	➢ OR	➢ Institutional **Repositories** in Developing Countries
➢ Without the words	➢ NOT	➢ Institutional **Affairs** of JNU

Table 3.9: Search and Display of Metadata Elements in DSpace at Vidyanidhi

Dublin Core Metadata Recorded in XML File	Browsable Metadata	Searchable Metadata	Displayed Metadata
➢ dc.title ➢ dc.creator ➢ dc.subject ➢ dc.description ➢ dc.publisher ➢ dc.contributor ➢ dc.date ➢ dc.type ➢ dc.identifier ➢ dc.language ➢ dc.coverage ➢ dc.rights ➢ thesis.degree	➢ Communities & Collections (Collection by Subject / Collection by University) ➢ Titles ➢ Authors ➢ By Date	➢ Keyword ➢ Author ➢ Title ➢ Subject ➢ Abstract	➢ Title ➢ Authors ➢ Keywords ➢ Issue Date ➢ Abstract ➢ URI ➢ Appears in Collections (Name of Collections)

3.4.3 Digitization Facility at Vidyanidhi Project

Although Vidyanidhi has a centralized digitization facility and infrastructure, some of its partner universities do not have any such facility. Thus, the collection of Vidyanidhi grows only when it shares its physical resources with other universities, which is not practically feasible in a vast country like India. Till date, most of the universities in India do not insist that researchers submit electronic copies of doctoral theses, either on CD-ROM or online. If electronic copies of doctoral theses are made available with the partner universities, they can proceed to upload e-theses into the Vidyanidhi ETD server.

3.4.4 Some Observations

Vidyanidhi is one of the few successful digital library projects in India having a clear roadmap. Over the time Vidyanidhi proved to be an exemplary success story from India that brings worldwide audience closer to Indian scholarly literature. However, there is a mismatch between the number of bibliographic records available in Vidyanidhi's Indian Theses Database and the number of full-text e-theses available in DSpace@Vidyanidhi portal. Indian Theses Database has more coverage than the DSpace@Vidyanidhi. Only a fraction of theses recorded in Indian Theses Database is available at DSpace@Vidyanidhi. This may be because Indian universities are not very keen to make their theses and dissertations available open access.

Sustainability is another major concern in this effort. However, host university of Vidyanidhi project has already tried to address the long-term sustainability issue involving different stakeholders and funding agencies.

3.5 Digital Repository of Theses and Dissertations of IISc [http://etd.ncsi.iisc.ernet.in]

The Digital Repository of Theses and Dissertations of Indian Institute of Science Bangalore (ETD@IISc) is an internationally acclaimed open access ETD repository of IISc and a joint information service of the National Centre for Science Information (NCSI) and the IISc Library.

This repository disseminates and preserves high quality research theses of the Indian Institute of Science. This service is a complement to ePrints@IISc, the research publications repository of IISc as shown in Figure 3.11.

Apart from electronic copies of recently submitted theses, this repository also includes digitized theses, where digitization of old documents is an on-going project at IISc library. Figure 3.12 depicts workflow of digital archiving process at ETD@IISc. This Figure indicates that a registered student needs a unique username and password to proceed to the self-archiving process. After login, the

student first submits metadata information related to his thesis using appropriate metadata entry form. Then he uploads his full-text thesis and related files in appropriate file formats. After completion of submission process, a record of the thesis work is stored in the buffer space of the repository. The submitted record is validated and approved by ETD coordinator and/or his thesis guide, and then record of thesis gets permanently archived in the ETD repository for worldwide dissemination.

Table 3.10 shows the subject-wise distribution of full-text theses in the repository which depicts that the Division of Mechanical Sciences has the maximum contribution to this open access archive, followed by the Division of Electrical Sciences and the Division of Biological Sciences. This Table also indicates that collection from Division of Physical and Mathematical Sciences has the maximum growth rate, followed by the Division of Chemical Sciences and the Division of Mechanical Sciences, as calculated within a gap of fourteen months. Figure 3.13 shows graphically the growth pattern of electronic theses collections at ETD@IISc.

Figure 3.14 depicts statistical distribution of electronic theses collections in ETD@IISc by subject divisions that shows the Division of Mechanical Sciences is the top contributor (35.7%), followed by the Division of Electrical Sciences (24.2%) and the Division of Biological Sciences (15.8%).

This repository uses a standard set of metadata elements as available in DSpace software such as Title, Authors, Subject, Keywords, Date of Submission, Publisher, Abstract, URI (Uniform Resource Identifier), and Name of Collection. Here the ‘author’ refers to both researcher (who carried out a thesis work) as well as supervisors. The record display page of individual thesis provides links to full-text objects, along with metadata information. Table 3.11 shows metadata elements used at ETD@IISc such as browsable metadata elements, searchable metadata elements and displayed metadata elements. This Table also indicates that ETD@IISc collections can be navigated by name of author, name of guide, thesis title, subject and issue date.

Indian Institute of Science established an 'IISc INDEST User Group' that has subgroups on 'Thesis Format Guidelines', 'Copyright Issues' and 'Technology Issues for Archival and Online Hosting'. ETD@IISc hosts some operational guidelines and agreement forms, derived from the recommendations of the User Group, for the perusal of IISc researchers and members of INDEST Consortium. This portal also provides *IISc Theses Style Templates* for wide use by IISc students and research community. Over the time this ETD initiative has became a role model amongst the members of INDEST Consortium. The INDEST Consortium has a strong advocacy role in formulation of policy for establishment of open access ETD repositories and/or open access institutional repositories in its respective member institutions.

The ETD@IISc is a flagship initiative for increasing visibility and availability of research literature produced in India's elite institutions. This happens to be a motivational factor for other Indian and Asian research institutions as well as to the research funding agencies to improve research landscape of emerging knowledge economies. This effort will obviously attract more talents and research funding into higher educational institutions in India.

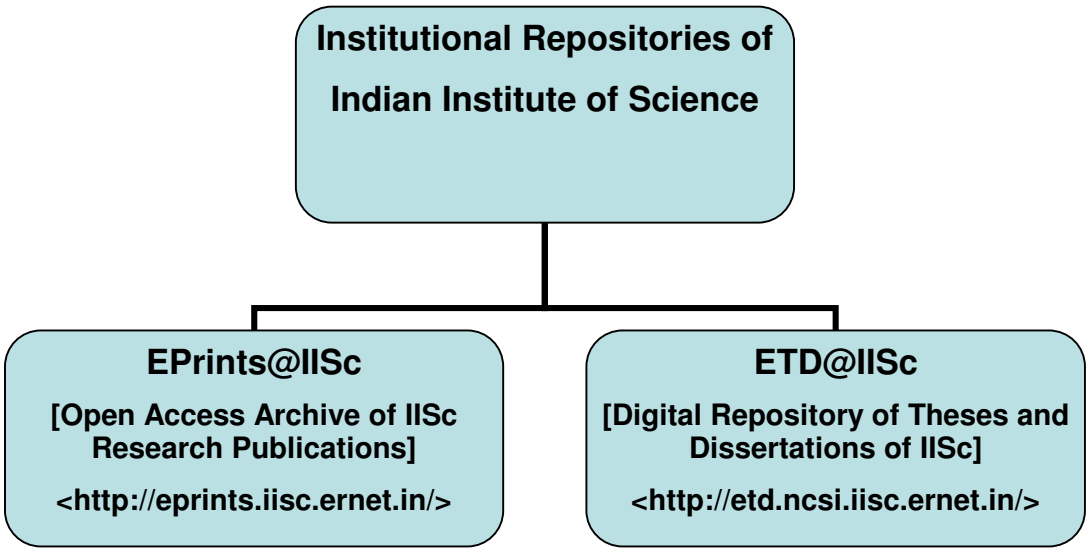


Fig. 3.11: Open Access Initiatives of Indian Institute of Science (IISc)

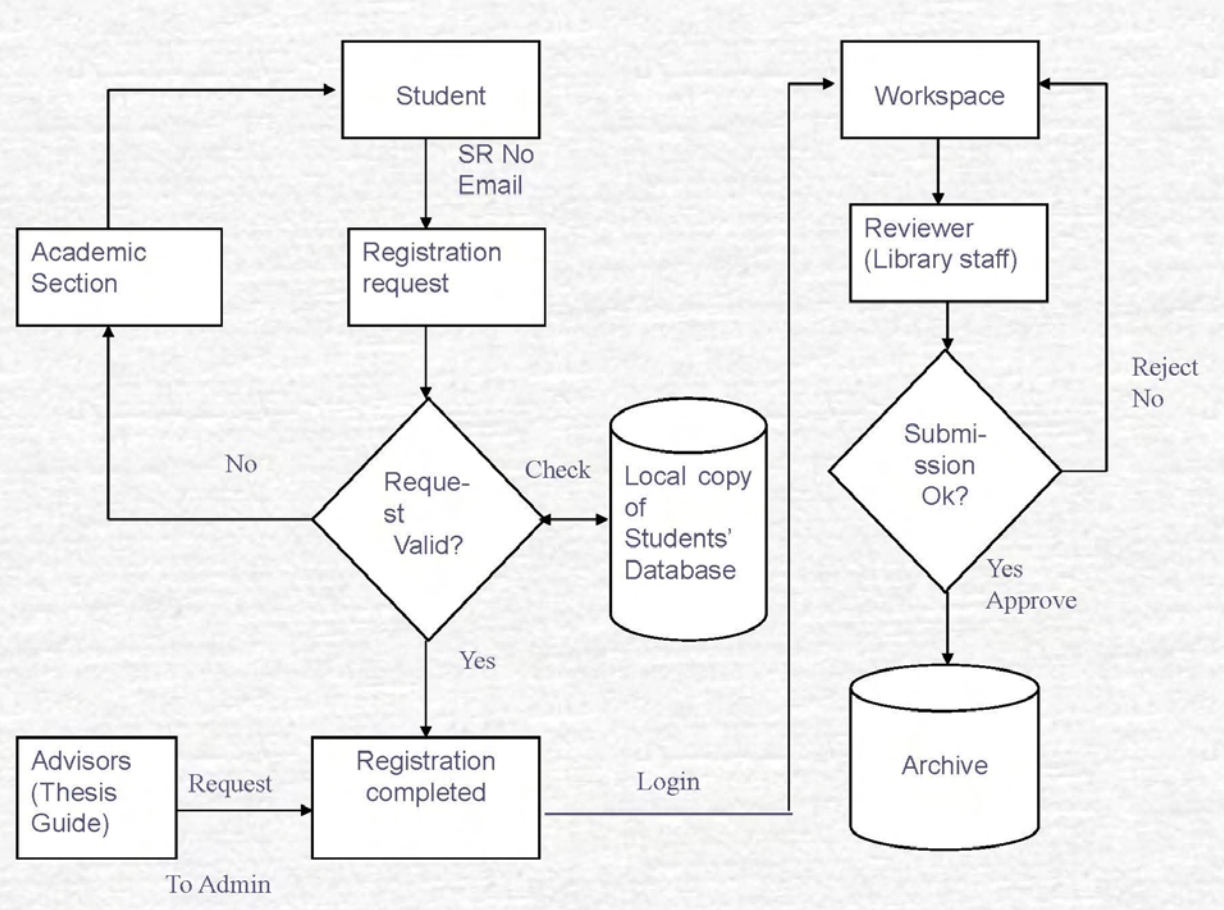


Fig. 3.12: Workflow of Digital Archiving Process in ETD@IISc

Table 3.10: Distribution of Electronic Theses at ETD@IISc

Name of Research Division	As on 21/07/2008	As on 15/05/2007	Growth Rate
Division of Biological Sciences	47	40	17.5
Division of Chemical Sciences	42	25	68.0
Division of Electrical Sciences	72	62	16.1
Division of Information Sciences	5	12	-58.3
Division of Mechanical Sciences	106	77	37.7
Division of Physical and Mathematical Sciences	25	13	92.3
Total	**297**	**229**	29.7

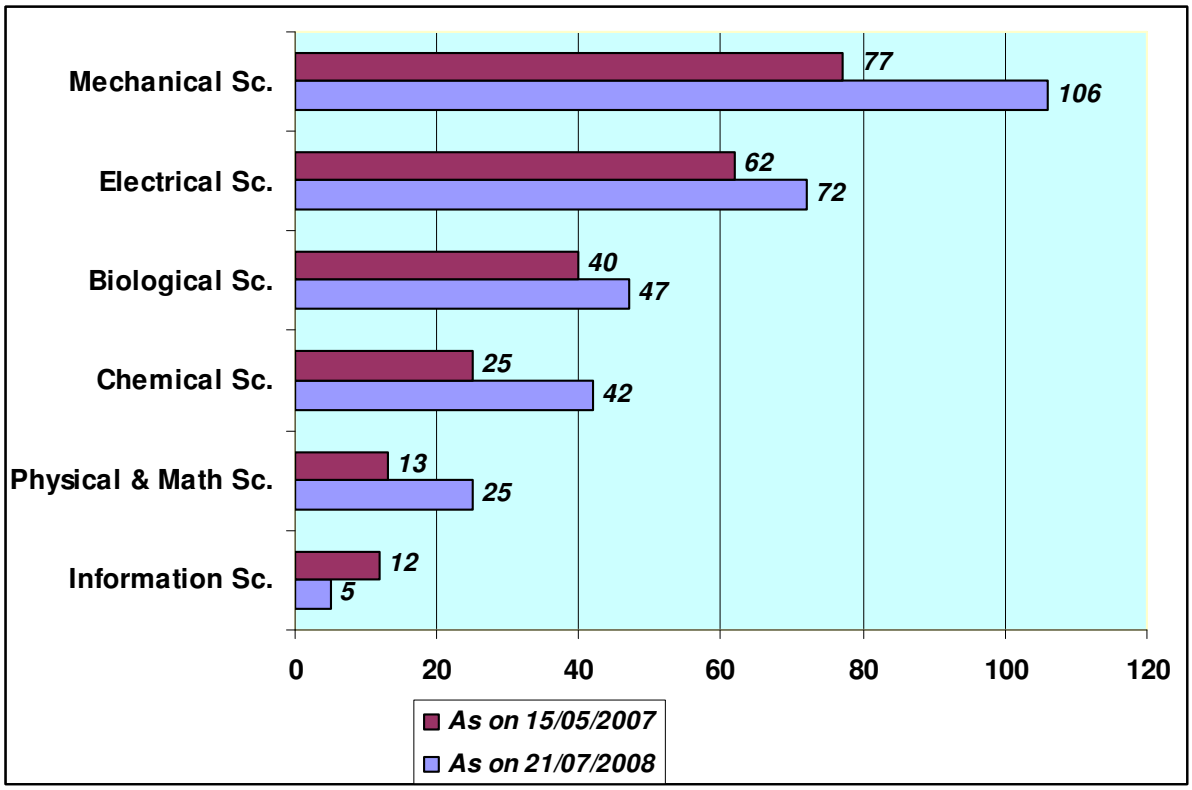


Fig. 3.13: Growth Pattern of Electronic Theses at ETD@IISc

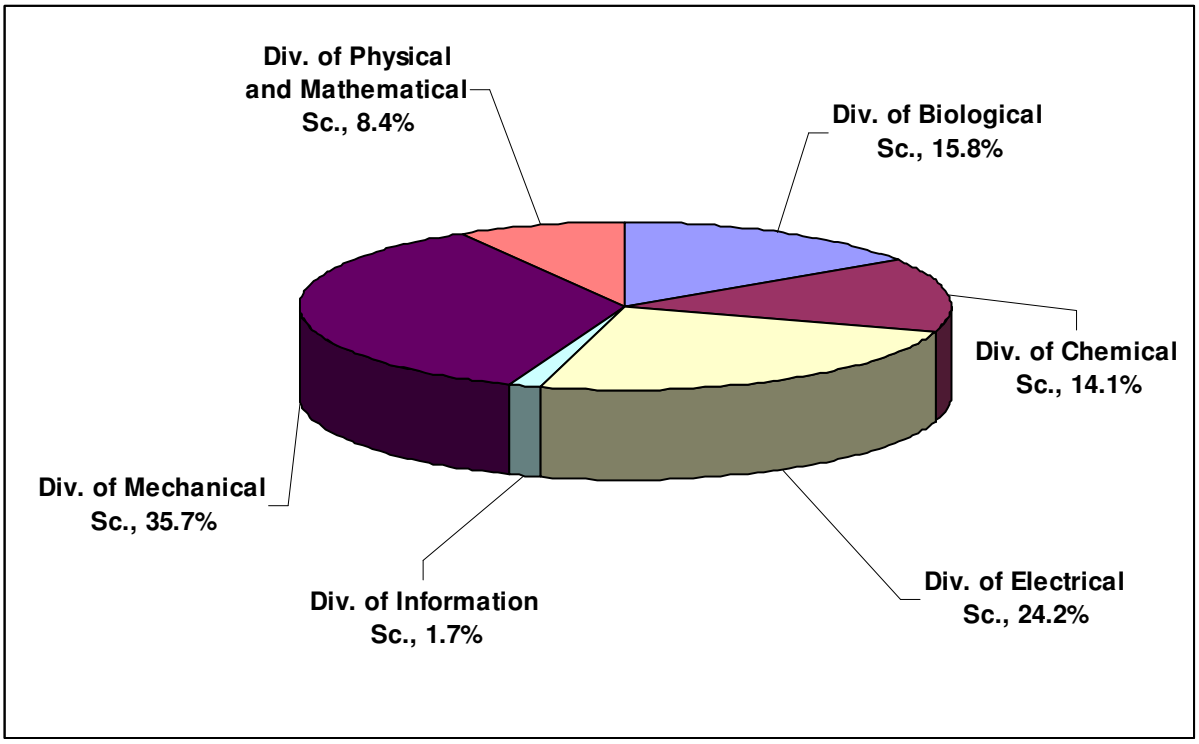


Fig. 3.14: Distribution of Electronic Theses in ETD@IISc by Subject Division

Table 3.11: Metadata Elements Used at ETD@IISc

Browsable Metadata	Searchable Metadata	Displayed Metadata
➢ Communities & Collections (Name of Collections) ➢ Titles ➢ Authors ➢ Thesis Guides ➢ Subjects ➢ Issue Date	➢ Keyword ➢ Author ➢ Title ➢ Subject ➢ Abstract	➢ Title ➢ Authors ➢ Issue Date ➢ Publisher ➢ Abstract ➢ URI ➢ Appears in Collections (Name of Collections)

3.6 OpenMed@NIC [http://openmed.nic.in]

The OpenMED@NIC is a national level open access archive for biomedical and allied sciences providing a self-archiving facility to biomedical researchers, practitioners and professionals. OpenMED is one of the flagship open access initiatives of Bibliographic Informatics Division of National Informatics Centre. Other important initiatives of this Division of NIC are: (i) IndMED: a bibliographic database of Indian biomedical research <http://indmed.nic.in>; and (ii) medIND: Indian full-text bio-medical journals indexed in IndMED <http://medind.nic.in>. Figure 3.15 depicts the biomedical open access information products of National Informatics Centre, where OpenMED serves as national level open access archive for biomedical disciplines.

The objective of this initiative is to encourage a self-archiving culture amongst biomedical researchers, practitioners and professionals in India so that cases in evidence-based clinical practices and biomedical research reports can be widely disseminated. The goal of this repository is to preserve valuable research publications for future biomedical researchers and publicize research being conducted in the country. Figure 3.16 illustrates objectives, goal and aim of the OpenMed initiative.

Although this repository has a wide coverage in all major biomedical subject areas, it has a comprehensive collection of papers on tuberculosis as well as tropical diseases.

Using this facility, authors/researchers/practitioners can self-archive their scientific and technical documents including theses and dissertations. They usually register once in order to obtain a unique user identity for the system. No registration is however required for searching the archive or viewing the documents. OpenMED accepts both published and unpublished documents having relevance to research in biomedical and allied sciences including paramedical areas. Self-archived contributions can be in the form of theses, dissertations, journal articles, conference papers, conference proceedings, posters, presentations, clinical cases, technical reports, working papers, and policy documents. This repository also accepts documents in Indian languages, but in this case descriptive metadata is entered in English. The self-archived document then goes to the moderator and becomes part of the OpenMED archive on its acceptance. Presently, this archive also serves as a knowledge repository of journals, institutions and conference organizers in India for long-term preservation of research publications. These include:

- Institutional Repositories
  - College of Veterinary Science and Animal Husbandry, Anand Agricultural University, Gujarat
  - National Tuberculosis Institute, Karnataka
  - Sir Ganga Ram Hospital, New Delhi
- Journal Repositories
  - Anatolian Journal of Psychiatry
  - Archives of Medical Science
  - Calicut Medical Journal
  - Indian Journal of Medical Informatics
  - Indian Journal of Tuberculosis
  - Indian Pacing and Electrophysiology Journal

- Indian Veterinary Journal
- JK Science
- Journal of Applied Basic Medical Sciences
- Journal of Indian Association for Child and Adolescent Mental Health
- Online Journal of Health and Allied Sciences
- OPUS 12 Scientist

This repository is compliant to OAI-PMH (Protocol for Metadata Harvesting) and indexed in the CASSIR (Cross Archive Search Service for Indian Repositories), Google Scholar, OAIster and OpenDOAR (Directory of Open Access Repositories).

Figure 3.17 depicts participatory collection development in OpenMED in partnership with self-archiving authors, biomedical researchers, conference organizers, biomedical journal publishers and biomedical institutions in India.

Figure 3.18 indicates that OpenMED stores a major collection from Indian journal publishers (57.2% of overall collection), followed by biomedical institutions (26.5%) as on 1 August 2008. It has a tiny collection of theses and dissertations (2.4% of overall collection) voluntarily self-archived by biomedical researchers in India.

Table 3.12 shows metadata elements used at OpenMED such as browsable metadata elements, searchable metadata elements and displayed metadata elements. This Table also indicates that OpenMED collections can be navigated by subject, year and type of repository.

OpenMED is a successful open access initiative in India that was nominated for many national and international awards in e-Health category. OpenMED was nominated as a finalist for the prestigious Stockholm Challenge 2008 Award.

Indian biomedical research funding agencies such as Indian Council of Medical Research (ICMR), Department of Biotechnology (DBT), Department of Science and Technology (DST), and Council of Scientific and Industrial Research (CSIR) provide research supports to the junior and senior research fellows (JRFs/ SRFs)

to carry out their doctoral and pre-doctoral research degrees. These funding agencies can make use of OpenMED to enhance international visibility of Indian theses literature. They can motivate Indian researchers, who are the recipients of research fellowships, to self-archive biomedical theses and dissertations into OpenMED portal.

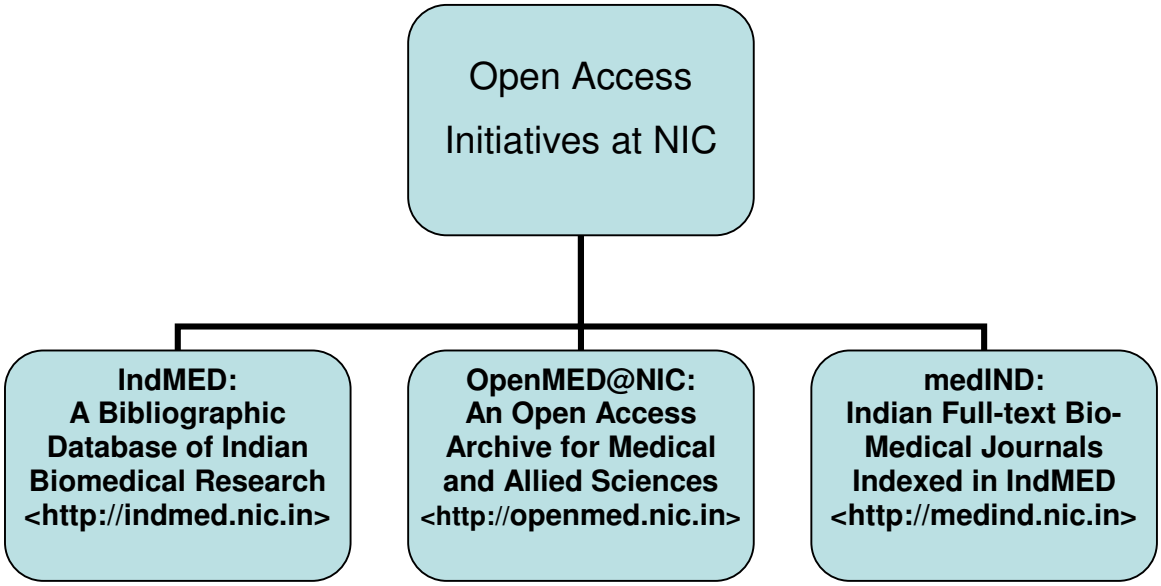
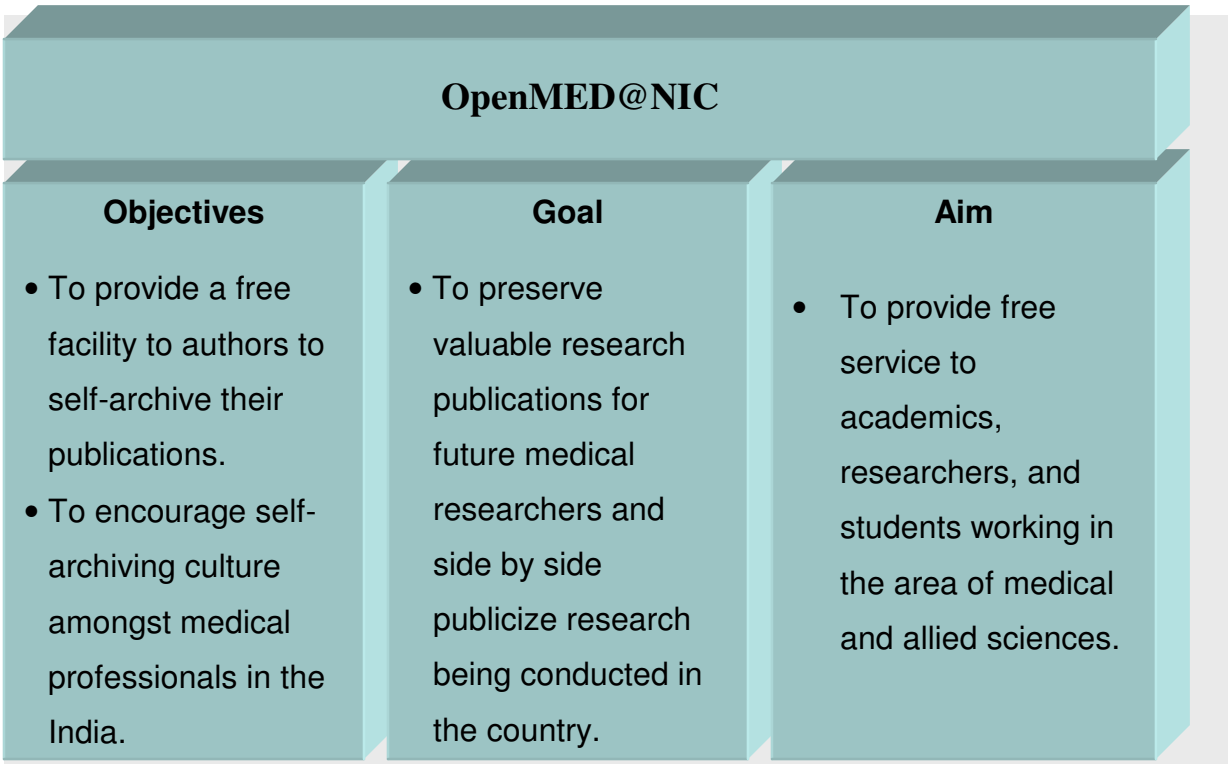


Fig. 3.15: Biomedical Information Products of National Informatics Centre



Source: *About OpenMED@NIC http://openmed.nic.in/information.html*

Fig. 3.16: Objectives, Goal and Aim of OpenMed Project

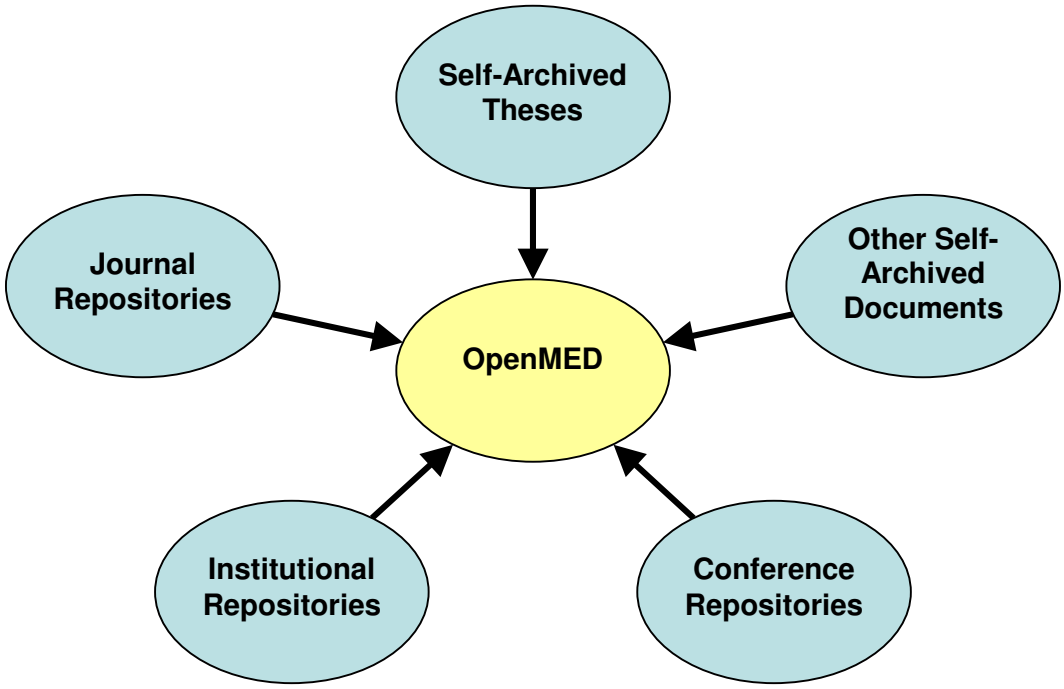


Fig. 3.17: Collection Development in OpenMed Portal

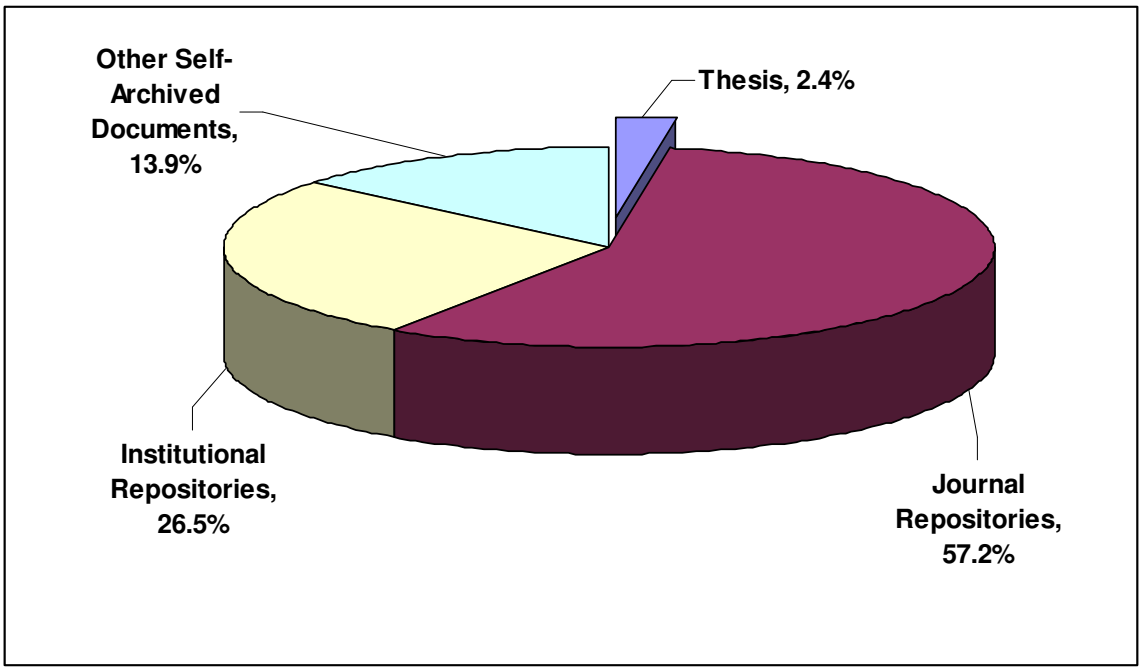


Fig. 3.18: Distribution of Different Types of Documents in OpenMED

Table 3.12: Metadata Elements Used at OpenMED

Browsable Metadata	**Searchable Metadata**	**Displayed Metadata (for Thesis)**
➢ Subject ➢ Year ➢ Type of Repository	➢ Title ➢ Author/ Creator ➢ Subject ➢ Date ➢ Abstract/ Full Text ➢ Keyword ➢ Language ➢ Document Type	➢ Title ➢ Author/ Creator ➢ Year ➢ University Name ➢ Abstract ➢ EPrint Type: Thesis (PhD) ➢ Additional Information ➢ Uncontrolled Keywords ➢ Subjects ➢ ID Code ➢ Deposited By

		➢ Date of Deposit

3.7 DSIR-supported Other ETD Initiative

The Technology Information Facilitation Programme (TIFP) of the Department of Scientific & Industrial Research (DSIR), Government of India, has a series of focused activities to support digital and indigenous knowledge base, and more particularly Indian Digital Library of Theses and R&D Publications. This activity aims at "establishing suitable mechanism for capturing full-text information of dissertations/theses and R&D reports of Government funded projects and publishing such material the same through electronic media" (DSIR, 2006).

Under the auspices of this TIF Programme, DSIR is supporting a project entitled "Design and Development of Database and Web-Portal of Indian Theses Covering Technology and Management aspects related to Manufacturing", undertaken by Department of Mechanical Engineering in the Bannari Amman Institute of Technology at Sathyamangalam, Tamil Nadu (DSIR, 2006).

The objectives of the project are:

- To develop a Web-based information platform for improving the quality of higher education in the fields of manufacturing technology and management in India;
- To promote research in various other areas in the fields of manufacturing technology and management by provision of ready access to information; and
- To review the practices for dissemination of scholarly information, which have developed in the last decade (1995 to 2005).

This project was initiated in 2005-06 financial year under the heading 'Indian Digital Library of Theses and R&D Publications' and is scheduled to be concluded in 2008-09 financial year. Unfortunately no Web-portal is launched till date and no significant progress on the project has been reported in journals or conferences.

On the other hand, the predecessor of this DSIR-supported ETD project - Vidyanidhi, is a grand success as University of Mysore had a clear roadmap and

strategies to achieve project objectives. Another contemporary DSIR-supported project 'Development of OAI-Based Institutional Research Repository Services in India', being implemented by National Centre for Science Information (NCSI) of Indian Institute of Science, Bangalore is also a successful initiative that aims "to facilitate improved access, visibility and impact of Indian science research output through the establishment of a network of interoperable, open access digital research repositories and related services in the country" (DSIR, 2006). An important outcome of the NCSI project 'Development of OAI-Based Institutional Research Repository Services in India' is described in following section.

3.8 Cross Archive Search Service for Indian Repositories (CASSIR) [http://ardb4.ncsi.iisc.ernet.in/oai/]

The Cross Archive Search Service for Indian Repositories (CASSIR) is a metadata harvesting service for Indian open access repositories. This Web-based search and browse service is a part of the ongoing project entitled 'Development of OAI-Based Institutional Research Repository Services in India', initiated by National Centre for Science Information (NCSI) of Indian Institute of Science (IISc) and supported by Department of Scientific and Industrial Research (DSIR). In this service, an OAI-PMH compliant software (PKP Harvester) harvests metadata from the registered open access repositories in India. At present, CASSIR covers eighteen Indian open access repositories such as:

- Catalysis Database (ePrints@NCCR IIT Madras)
- DRS at National Institute of Oceanography
- DSpace at Icfai Business School (IBS), Ahmedabad
- IIMK's Scholarship Repository
- DSpace at National Chemical Laboratory, Pune
- Vidyanidhi Digital Library
- Librarians' Digital Library
- Dspace@NIT Rourkela
- DU Eprint Archive

- ePrints@IISc
- ePrints@IIT Delhi
- ETD@IISc
- Open Access Repository
- DSpace @ ISI Bangalore
- NAL Institutional Repository
- OneWorld South Asia Open Archive Initiative
- OpenMED
- RRI Digital Repository

Some of these open access repositories store a number of full-text e-theses. Thus, CASSIR serves a functionality of cross-searching of repositories. This cross-searching functionality is much needed in a country like India where hundreds of open access repositories are expected to be activated in near future, if the *UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format)* regulations can be implemented soon. Figure 3.19 shows the advanced search interface of CASSIR, where searching of ETDs in Indian repositories can be carried out easily by limiting document type as 'thesis'.

Table 3.13 shows metadata elements used at CASSIR such as browsable metadata element, searchable metadata elements and displayed metadata elements. This Table also indicates that CASSIR collections can be searched by range of years, name(s) of repository, type of document, or any other metadata element. CASSIR collections can only be navigated by name of repository.

With the proliferation of Indian open access repositories and expected proliferation of Indian ETD repositories, CASSIR stands a unique cross-searching platform in order to make it convenient for end-users.

Table 3.13: Metadata Elements Used at CASSIR

Browsable Metadata	Searchable Metadata	Displayed Metadata
➢ Name of Archive	➢ Name of Archive ➢ Title ➢ Creator ➢ Date – From & Until ➢ Language ➢ Identifier ➢ Description ➢ Format ➢ Publisher ➢ Rights ➢ Type ➢ Subject ➢ Relation ➢ Contributor ➢ Source ➢ Coverage	➢ Creator ➢ Contributor ➢ Date ➢ Identifier (Hyperlink) ➢ Description ➢ Format ➢ Language ➢ Rights ➢ Subject ➢ Title ➢ Type ➢ Name of Archive

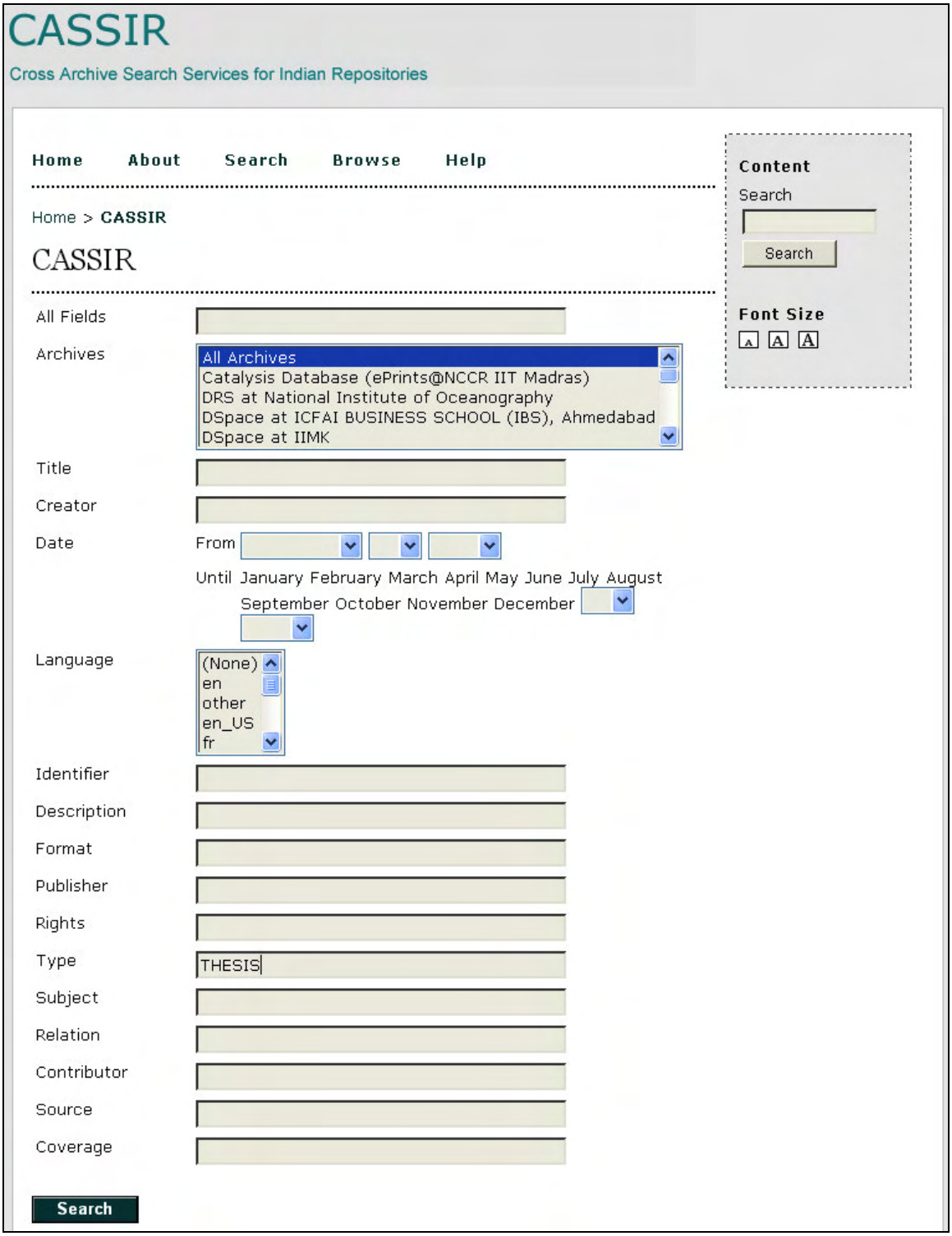


Fig. 3.19: Advanced Search Interface of CASSIR

3.9 Online Bibliographic Databases of Theses and Dissertations

In India some efforts are being made for bibliographic control of Indian theses and dissertations, particularly doctoral theses. UGC has also proposed creation of Indian National Theses Database (INTED), which is an online centralized bibliographic database for Indian theses and dissertations. The INTED was conceptualized in the "UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005". Few other national online bibliographic databases of theses and dissertations are also in existence for some years which can be considered as predecessors of INTED. These online databases have a standard set of metadata elements, with or without abstracts, to inform end-users about the details of theses and dissertations awarded in Indian institutions and universities. Sometimes, the institutions granting academic degrees also maintain online bibliographic databases of theses and dissertations. Table 3.14 provides an indicative list of online bibliographic databases of theses and dissertations in India at the national or institutional level.

Association of Indian Universities (AIU) also maintains an offline bibliographic database of doctoral theses awarded in Indian universities. AIU brings out a list of new additions to this database on weekly basis in the *University News* (an official journal of AIU). AIU published a series of "Bibliography of Doctoral Dissertations" accepted by Indian universities in the subject areas of humanities, social sciences and sciences.

In the following sections, these bibliographic databases are analyzed along with a brief introduction.

Table 3.14: Online Bibliographic Databases of Theses and Dissertations in India

Name or the Repository	Name of Institution	URL	Level
Vidyanidhi	University of Mysore	www.vidyanidhi.org.in	National
CSIR e-Thesis	URDIP	http://csirexplorations.com/ethesis/e_thesis.htm	National
National Online Union Catalogue of Doctoral Theses	INFLIBNET Centre	http://unicat.inflibnet.ac.in:8080/unicat/thesis.jsp	National
DELNET Theses and Dissertations Database	Developing Libraries Network (DELNET)	www.delnet.nic.in	National
ETD Digital Library @ IIT Bombay	Indian Institute of Technology Bombay	http://www.library.iitb.ac.in/~mnj/gsdl/cgi-bin/library	Institutional
PhD Theses @ IIT Madras	Indian Institute of Technology Madras	www.cenlib.iitm.ac.in/docs/library/index.php?page=theses	Institutional

3.9.1 CSIR e-Thesis [http://csirexplorations.com/ethesis/e_thesis.htm]

The CSIR e-Thesis is an online bibliographic database of doctoral theses submitted by CSIR research fellows. This is an initiative of CSIR Unit for Research and Development of Information Products (URDIP) at Pune, India. E-Thesis database is an authoritative source for information about doctoral theses and dissertations, submitted in Indian universities by the researchers, who availed CSIR Junior Research Fellowship (JRF) or Senior Research Fellowship (SRF) scheme. CSIR research fellows carry out their research work in over 250 national laboratories, universities and research institutes in India. This database is popularly used by doctoral candidates, faculty members and scientists who would like to know the topics of already completed researches by other research fellows in past.

At the time of submission of PhD thesis to an Indian university, a CSIR research fellow usually submits electronic copy of his/her thesis on CD-ROM to URDIP along with a descriptive set of metadata for inclusion in this online database. Figure 3.20 shows contribution pattern as followed in the CSIR e-Thesis database. This Figure depicts that CSIR research fellows attached with CSIR laboratories or Indian universities are the main contributors, whereas contribution from CSIR staff members can also be accepted in this database.

The research fellow receives a token financial incentive (INR 3000.00, about US Dollar 73.00) to recover the cost of creation of CD-ROM. As CSIR is the largest supporter of doctoral research work in the country, URDIP in association with Human Resources Development Group (HRDG) of CSIR is creating a digital library, containing CSIR supported PhD Theses in the 11th Five Year Plan period (2007-2012) of the country. Probably, the existing CSIR e-Thesis database will be scaled up for this purpose.

CSIR e-Thesis database has different searching and browsing options such as browse, simple search, advanced search and thesis bibliography. This database can be searched by applying different parameters using Boolean operators such as Title, Author, Guide, Institute, University, Synopsis, and All Fields. The

retrieved record of a thesis is displayed with the metadata such as e-Thesis Number, Title, Author, Position, Guide, University, Institute, Area, Specialization, Submitted Year, Synopsis, Papers (hyperlinks), and Patents (hyperlinks). Table 3.15 depicts the browsable, searchable and displayed metadata elements in CSIR e-Thesis database.

Table 3.16 and Figure 3.21 illustrate the subject-wise distribution of this bibliographic database. This Table indicates that the largest number of theses belongs to chemistry, which suggests that CSIR gives more priority to chemical sciences and chemical engineering than other S&T areas. This Table also suggests that the e-Thesis database probably does not have comprehensive coverage, but created bibliographic records only from few CSIR institutions.

Table 3.17 indicates a detailed template and metadata schema for submission of thesis record in the CSIR e-Thesis database. This template is very comprehensive and meticulously followed by the CSIR research scholars at the time of submission of e-thesis on CD-ROM. Hyperlink to research papers and patents is a unique option that facilitates navigation of intellectual contributions of researcher in the forms of patents and research papers.

Table 3.15: Metadata Elements Used at CSIR e-Thesis

Browsable Metadata	Searchable Metadata	Displayed Metadata
➢ Subject ➢ Keyword (controlled)	➢ Title ➢ Author ➢ Guide ➢ Institute ➢ University ➢ Synopsis	➢ E-Thesis No. ➢ Title ➢ Author ➢ Position ➢ Guide ➢ University ➢ Institute ➢ Area ➢ Specialization ➢ Submitted Year ➢ Synopsis ➢ Papers (Hyperlink) ➢ Patents (Hyperlink)

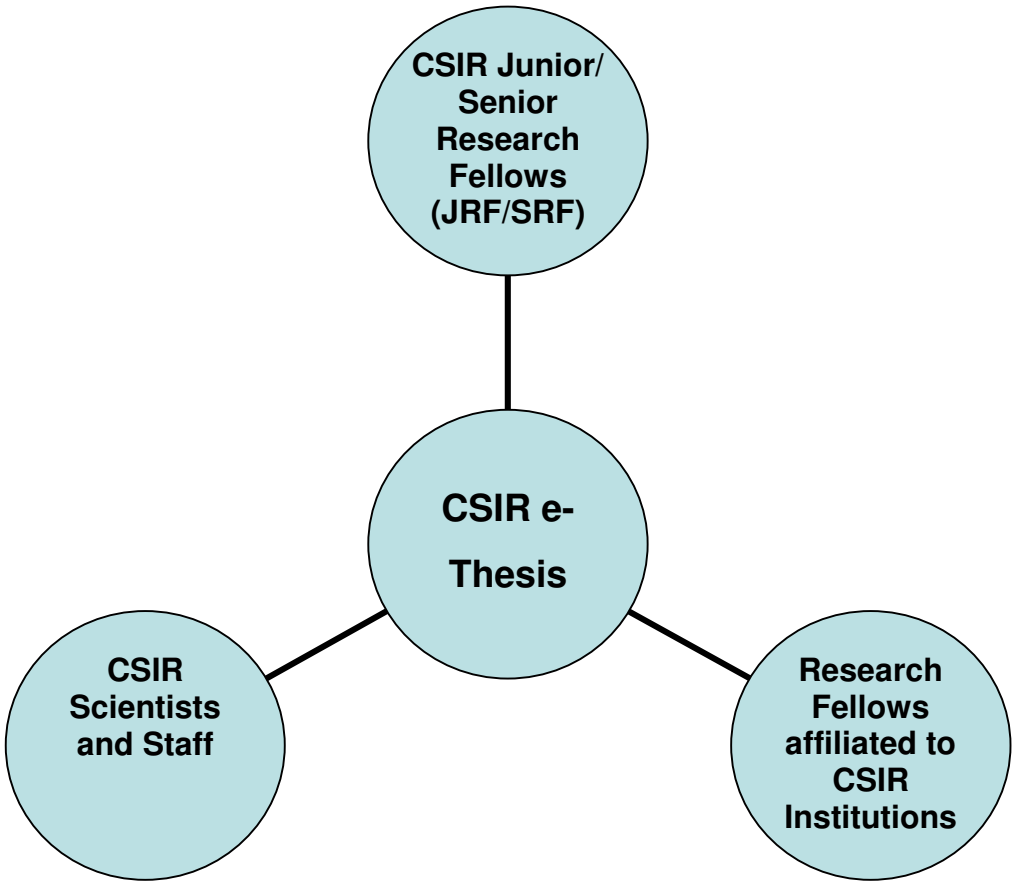


Fig. 3.20: Contribution Pattern in CSIR e-Thesis Database

Table 3.16: Subject-wise Distribution of CSIR e-Thesis Database

Area	**Number of Records** *(As on 22/07/08)*	**Coverage** (Percent)
Chemistry	542	82.25
Biology	54	8.19
Physics	40	6.07
Biotechnology	11	1.67
Pharmaceuticals	2	0.30
Engineering	2	0.30
Leather Processing	2	0.30
Botany	1	0.15
Earth Sciences	1	0.15
Environment	1	0.15
Medicine	1	0.15
Others	2	0.30
Total	**659**	**100.00**

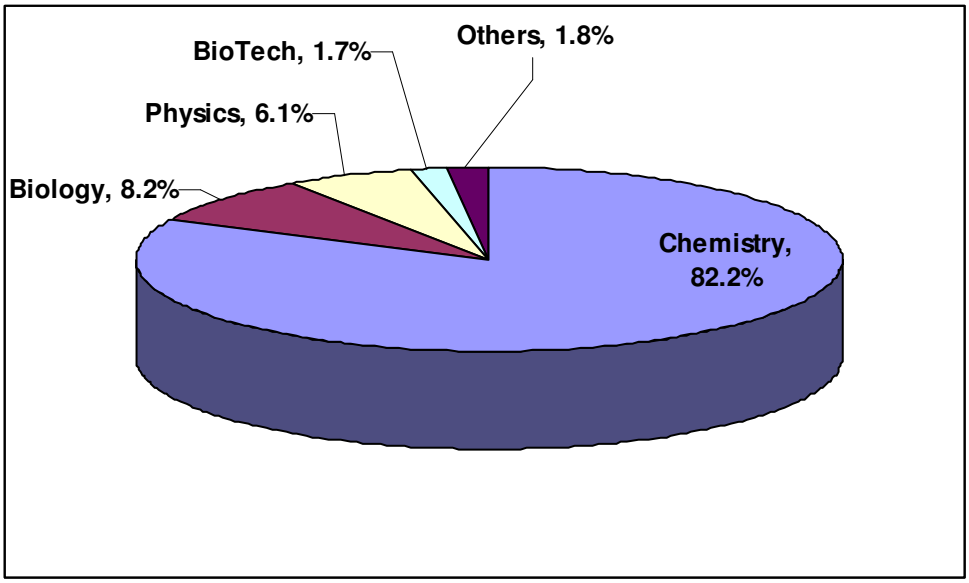


Fig. 3.21: Subject-wise Distribution of CSIR e-Thesis Database

Table 3.17: Metadata Template for Inclusion in CSIR e-Thesis Database

Metadata Element	Details of Metadata Element
Title	Title of the Thesis
Author Details	(a) Name of the research student (b) CSIR number awarded (c) The rank held at CSIR (JRF, SRF, RA, etc) (d) Contact e-mail (e) Qualifications (f) Address for correspondence (g) Permanent address
Guide Details	(a) Name of the guide (b) Contact e-mail (c) Qualifications (d) Address for correspondence (e) Permanent address
Degree	Degree awarded for the thesis
University	Name of the university where the thesis is submitted
Institute/Lab	Name of the CSIR lab / academic institution
Date of Submission	Date of submission of thesis to the university
No. of Pages	Total No. of pages of the document
Subject/Area	Research area (with discipline and specialization)
Abstract/Synopsis	Abstract appears here
Publications	Full bibliographic details of published papers
Patents	Full bibliographic details of patents obtained/applied pertaining to this thesis
Source: *Guidelines for Submission of Dissertations in Digital Format http://csirexplorations.com/ethesis/submit_thesis.htm*	

3.9.2 INFLIBNET's National Online Union Catalogue of Doctoral Theses

National Online Union Catalogue of Doctoral Theses (http://202.141.130.73/unicat/thesis.jsp) is a database of bibliographic records of doctoral theses submitted by more than 130 universities/institutes in India. This is an initiative of Information and Library Network (INFLIBNET) Centre of University Grants Commission (UGC). This database consists of over 1,70,000 unique records from all subject areas, contributed by participating libraries of Indian universities or collected from various reliable sources and university announcements. As this database is created by a specialized centre of UGC, probably this is a predecessor of proposed Indian National Theses Database (INTED).

This online union catalogue has interfaces of basic search and advanced search. The database can be searched by title of the thesis, name of the researcher, name of the guide, name of the university, place of publisher, year of award, subject descriptor, and a Boolean combination. The search result generates a display page, containing the metadata of title of thesis, name of creator/researcher, name of university, year of award, subject/keywords, language, along with details of library held this thesis (for inter-library loan request). Figure 3.22 depicts the basic search interface along with descriptions of metadata schema adopted by this union catalogue. Figure 3.23 shows how advanced search can be carried out using Boolean operators.

Indian universities are familiar with this National Online Union Catalogue of Doctoral Theses. A union catalogue is a collaborative work, crafted with active community participation. If UGC or its parent body Ministry of Human Resources Development (MHRD) supports digitization initiatives for theses and dissertations in Indian universities in near future, then this Union Catalogue will become a resource base. Therefore, INFLIBNET Centre can also become a leading national agency for centralized hosting of digital library of theses and dissertations.

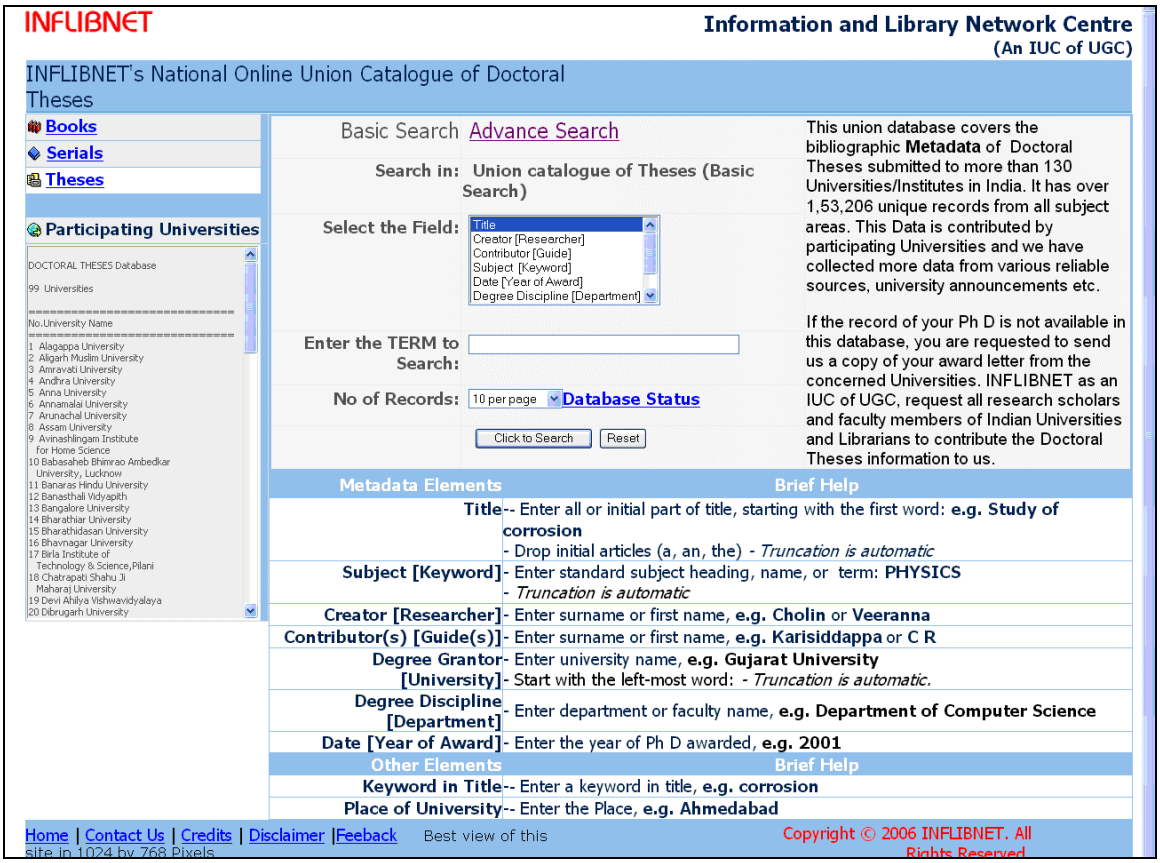


Fig. 3.22: Basic Search Interface of National Online Union Catalogue of Doctoral Theses

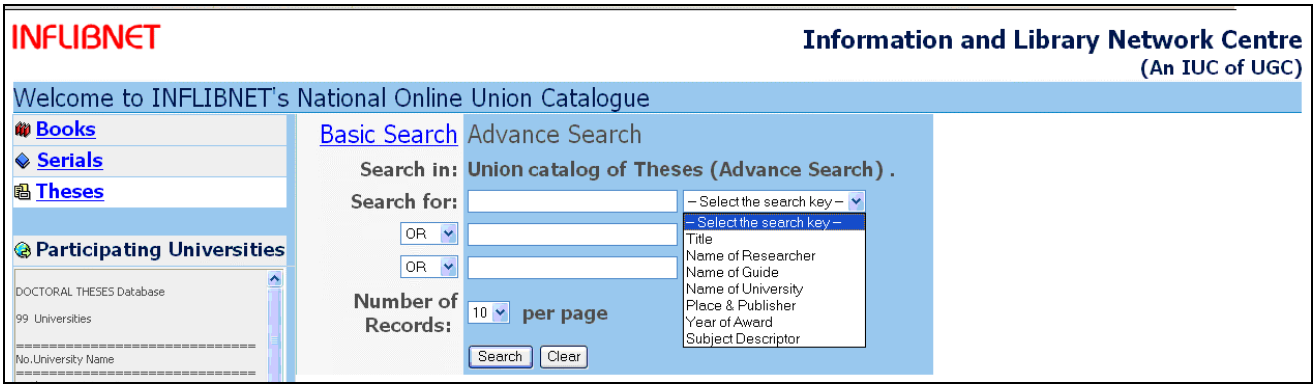


Fig. 3.23: Advanced Search Interface of National Online Union Catalogue of Doctoral Theses

3.9.3 DELNET Theses and Dissertations Database

DELNET Theses and Dissertations Database is an online union catalogue of bibliographic records of theses and dissertations available with member libraries (about 1100) of Developing Library Network (DELNET) (http://delnet.nic.in). This union catalogue is a closed access database, accessible to the library users of member institutions only. This database contains about 45,000 records of theses and dissertations. The search interfaces of this database generate a record display page with metadata such as title of thesis, author, subject descriptors, year of submission, location, physical document, type of material, place and publisher, etc. This database contains records of master's and higher level theses and dissertations, including project reports of post-graduate courses. This database is merely a union catalogue of member institutions, and cannot be considered as an authoritative database of theses and dissertations.

3.10 Some Observations

This chapter indicates that national ETD initiatives are still at the nascent stage where more concrete action plans are needed for digitization of back volumes of theses and dissertations, held in Indian university libraries. The stakeholders are mostly aware of the implications of open access, but they are awaiting for some concrete policy frameworks by the national accredited and granting agencies. National capacity building and self-sufficiency in maintaining electronic theses and dissertations can only be achieved by concerted approaches to this problem from national policymaking agencies. The maintenance of quality and standards of research degrees is considered as a major hurdle in expansion of higher education system in India. Open access to theses literature will bring the research literature under the purview of critical studies by scholarly forums and public review. If the public-funded researches produce futile results, the critical forums can press the funding agencies to adopt a mechanism for quality checks. The open access literature can also emerge as a safeguard mechanism to reduce risks of duplication in research efforts. Hence, open access to theses literature in a developing country like India is need of the hours to produce social

goods and social applications, rather than merely providing research degrees to the beneficiaries.

3.11 References

Department of Scientific and Industrial Research (2006). *DSIR Annual Report 2005-2006*. New Delhi: Department of Scientific and Industrial Research.

Ghosh, S.B. & Das, A. K. (2007). Open Access and Institutional Repositories - a Developing Country Perspective: a Case Study of India. *IFLA Journal*, 33(3), 229-250.

INDEST Consortium (2003). *Minutes of the Meeting held on 7th October, 2003 at IIT Delhi*. Retrieved 15 May 2007 from http://paniit.iitd.ac.in/indest/extended/minutes7thoct.pdf.

National Knowledge Commission (2006). *Report to the Nation 2006*. New Delhi: National Knowledge Commission.

National Knowledge Commission (2007). *Report to the Nation 2007.* New Delhi: National Knowledge Commission.

Planning Commission (2006). *Report of the Working Group on CSIR - Eleventh Five Year Plan 2007-2012*. New Delhi: Planning Commission.

Planning Commission (2006). *Report of Working Group on Higher Education - 11th Five Year Plan*. New Delhi: Planning Commission.

University Grants Commission (2007). *Guidelines for Junior Research Fellowship in Sciences, Humanities and Social Sciences*. Retrieved 15 May 2007 from www.ugc.ac.in/financialsupport/guideline_10.html.

University Grants Commission (2005). *UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005*. Retrieved 15 May 2007 from http://www.ugc.ac.in/new_initiatives/etd_hb.pdf.

Venkadesan, S. & Rajashekar, T.B. (2003). *Operational Guidelines for INDEST Digital Collection of Dissertations in Science and Engineering*. Retrieved 15 May 2007 from http://paniit.iitd.ac.in/indest/extended/indestmeet_oct03.pdf.

Virginia Tech (2003). *Proceedings of the Indo-US Workshop on Open Digital Libraries and Interoperability*, held at Ballston, USA, 23-25June 2003;

organized by Virginia Tech & Old Dominion University, USA and University of Mysore & Indian Institute of Science, India. Retrieved 15 May 2007 from http://fox.cs.vt.edu/IndoUSdl/tabl.htm

-*-

CHAPTER 4

Digitization of Documentary Heritage Collections: Cases of Indira Gandhi National Centre for the Arts and National Mission for Manuscripts

The cultural wealth of the world is its diversity in dialogue.

Koïchiro Matsuura

Director-General, UNESCO

4.1 Introduction

India is a country where cultural diversity and cultural pluralism are coexisting for centuries. Over the time Indian cultural institutions became the repositories of rich collections of cultural heritage resources embracing culturally and linguistically diverse communities across states of India. While traditional knowledge of linguistically diverse communities is largely un-documented, there were several attempts to collate them. Systematic documentation of traditional knowledge is centuries old practice of scholars and researchers to make the knowledge re-usable by future generations. These documentation initiatives ended up with producing literature of various kinds. On the other hand, some of the documentary heritage resources available with Indian institutions are on the verge of extinction due to lack of preservation and conservation initiative at the institutional level.

As a member country of UNESCO, India became automatic signatory of the UNESCO Universal Declaration on Cultural Diversity, adopted unanimously by the General Conference at its 31st session held on 2 November 2001. This is an international standard-setting legal instrument which raises cultural diversity to the rank of "common heritage of humanity" (UNESCO, 2001). The Declaration attempts to respond to two major concerns: (i) to ensure respect for cultural identities with the participation of all peoples in a democratic framework, and (ii)

to contribute to the emergence of a favourable climate for the creativity of all, thereby making culture a factor of development.

Article 6 of the Declaration further states:
'Towards Access for All to Cultural Diversity':
"While ensuring the free flow of ideas by word and image care should be exercised that all cultures can express themselves and make themselves known. Freedom of expression, media pluralism, multilingualism, equal access to art and to scientific and technological knowledge, including in digital form, and the possibility for all cultures to have access to the means of expression and dissemination are the guarantees of cultural diversity."

This Article emphasizes on the equitable access to culturally diverse multilingual contents with help of digital technologies. Modern information and communication technologies (ICT) have tremendous potentials to act as an enabler for intercultural dialogue through digital dissemination of cultural information, particularly with culturally diverse contents. Cultural informatics can also bridge linguistically diverse contents through translations and adaptations. Thus, cultural informatics can help in making culture a factor of development.

UNESCO and its member states adopted an action plan for the implementation of the UNESCO Universal Declaration on Cultural Diversity. Some of main lines of the action plan embraced digital technologies for strengthening the access to diverse cultural resources available across the country. Text Box 4.1 provides a list of selected main lines of action plan related to access and dissemination, for implementation of the Declaration. As a member state of UNESCO, India is committed herself to take active part in the main lines of action plan for implementation of the Declaration. Several digitization and digital library projects, described in this Chapter, have created an appropriate atmosphere for intercultural dialogue and intercultural partnership. These public-funded initiatives have tried to digitally include different communities in India, thus, further enhancing scope of cultural diversity and cultural pluralism in India (MCIT, 2004).

Text Box 4.1: Selected Main Lines of Action Plan for the Implementation of the UNESCO Universal Declaration on Cultural Diversity

- Encouraging "digital literacy" and ensuring greater mastery of the new information and communication technologies, which should be seen both as educational discipline and as pedagogical tools capable of enhancing the effectiveness of educational services.
- Promoting linguistic diversity in cyberspace and encouraging universal access through the global network to all information in the public domain.
- Countering the digital divide, in close cooperation in relevant United Nations system organizations, by fostering access by the developing countries to the new technologies, by helping them to master information technologies and by facilitating the digital dissemination of endogenous cultural products and access by those countries to the educational, cultural and scientific digital resources available worldwide.
- Ensuring protection of copyright and related rights in the interest of the development of contemporary creativity and fair remuneration for creative work, while at the same time upholding a public right of access to culture, in accordance with Article 27 of the Universal Declaration of Human Rights.
- Ensuring protection of copyright and related rights in the interest of the development of contemporary creativity and fair remuneration for creative work, while at the same time upholding a public right of access to culture, in accordance with Article 27 of the Universal Declaration of Human Rights.

Source: *Universal Declaration on Cultural Diversity - Main Lines of an Action Plan <www.un-documents.net/udcd.htm#action>*

4.2 Indira Gandhi National Centre for the Arts (IGNCA)

The Indira Gandhi National Centre for the Arts (IGNCA) is the youngest and one of the vibrant institutions under the Ministry of Culture, Government of India. Established in 1987, IGNCA becomes an advance centre for research, with active role in documentation and dissemination in the field of the arts, and more particularly in the areas of Indian art, culture and heritage. IGNCA is a focal agency in India for the implementation of the UNESCO Universal Declaration on Cultural Diversity through a series of intervention programmes, projects and initiatives.

IGNCA maintains six functional units – namely (i) Kalanidhi, (ii) Kalakosa, (iii) Janapada Sampada, (iv) Kaladarsana, (v) Cultural Informatics Lab and (vi) Sutradhara. The Kalanidhi is a multimedia reference library having an array of print and non-print collections. Kalakosa is the publication division of IGNCA, devoted mainly to the study of fundamental texts, predominantly in Sanskrit. Janapada Sampada is the division engaged in lifestyle studies. Kaladarsana provides the venue and forum for an intercultural creative dialogue. Cultural Informatics Lab applies technology tools for digital preservation of artifacts and cultural expressions. Sutradhara is the administrative division of IGNCA (Gaur, 2006; Paul & Jain, 2008; Jha, 2005; Jha, 2006).

Figure 4.1 provides a glimpse of organizational structure of IGNCA. The first four divisions are academic wings of IGNCA engaged in advanced research and academic discourses. Cultural Informatics Lab is interlinked with other five divisions, providing necessary ICT infrastructural and technical support necessary for interlinking and real-time collaboration. Sutradhara provides strategic and organizational support to other five divisions to make the programmes and projects successful.

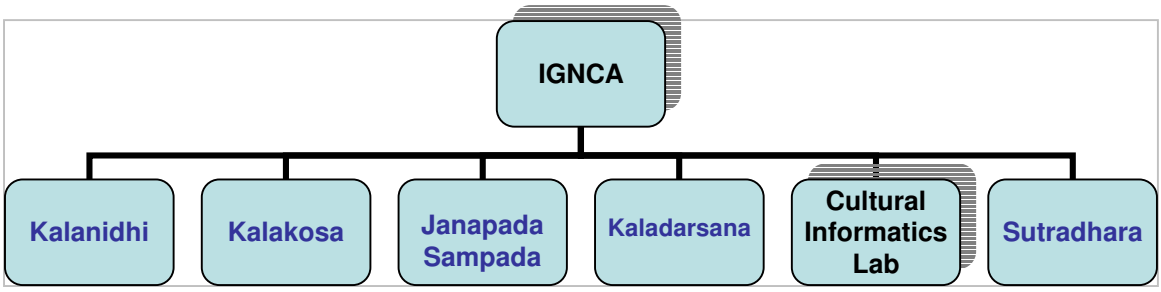


Fig. 4.1: Organizational Structure of IGNCA

4.2.1 Cultural Informatics Lab

Cultural Informatics Lab (CIL) is the finest world-class multimedia documentation centre in South Asia, having excellent ICT infrastructural facilities and digitization equipment. CIL, established in 1994, emerged from UNDP-funded capacity development project titled "Strengthening National Facility for Interactive Multimedia Documentation of Cultural Resources". As outcomes of this project, successfully completed in 2001, CIL enhanced its capacity, capability and technical expertise of multimedia documentation for multilingual contents. Now CIL is engaged in action research, particularly in the areas of interactive multimedia documentation and digitization of documentary heritage collections (IGNCA, 2007; Jha, 2005; Jha, 2006).

4.2.1.1 ICT Infrastructural Facilities

Over the time, CIL accumulated superior ICT infrastructure, starting from digitization workstations, to high speed network connectivity, high capacity storage system, state-of-the-art Web server, and other network resources. Figure 4.2 indicates some of the major ICT infrastructural facilities available at CIL. As CIL is a highly productive division of IGNCA, its ICT resources are growing significantly to accommodate newly created cultural contents.

4.2.1.2 Digitization Equipment and Facilities at CIL

CIL is one of the advanced digitization centres in south Asia, having digitization equipment and facilities for all kind of documents – both print and non-print ones. CIL has equipment for digitization of manuscripts, rare books, colour books, microfilms, microfiches, and slides. Scanned textual images can be converted into editable digital texts through the process of optical character recognition (OCR). CIL has necessary OCR software to convert scanned documents available in English or some other European scripts into editable digital texts.

CIL has facilities to covert analogue audios stored in audio tapes and gramophone records into digital audio file format, e.g. .WAV, MP3. Similarly, CIL has facilities to covert analogue videos stored in video tapes and motion picture films into digital video file format, i.e., MPEG 1 or MPEG 2.

In addition to digitization hardware, all necessary software for scanning all kinds of materials are available with CIL. Figure 4.3 indicates a range of scanners and other digitization equipment available with CIL.

4.2.1.3 Digitization Activities at CIL

Equipped with excellent infrastructural facilities and equipment, CIL offers all-round digitization solutions to its partner institutions and clienteles. Any kind of analogue documents can be converted into digital objects at IGNCA premises. Digitization activities carried out in CIL are mostly in-house. Its scanners and other equipment are mostly placed at IGNCA premises, to contribute to different digital library projects with centralized digitization facilities. CIL undertakes digitization of many kinds of materials such as manuscripts, archival records, rare books, rare photographs, back volumes of periodicals, photographic slides, microfiches, microfilms, audio tapes and video tapes. Thus, CIL offers long-term digital preservation services for multimedia and multilingual contents without compromising the output quality. Figure 4.4 indicates a range of scanning and digitization services offered by CIL to its partner institutions.

Over the time, CIL becomes a focal digitization and interactive multimedia documentation centre for institutions supported by Ministry of Culture. Ministry of Communication and Information Technology also recognized CIL as a dedicated national centre for cultural informatics, capable to handle large-scale digitization projects in documenting Indian cultural heritage and long-term preservation of precious cultural materials on the verge of extinction.

While centuries old priceless manuscripts, paintings, archival records and rare books available with state-own memory institutions are slightly better than those artifacts available with private collections, their nature of brittleness deters the digitization process to be undertaken by novice technicians at vendors locations. Thus, there is a need of super specialty centre that will have sophisticated technologies and technical expertise to handle rare materials with extreme care. In one decade, CIL became most sought-after national digitization centre for documentary artifacts and antiquities.

Having technical expertise in handling digitization of visual images, CIL developed some online utility tools for viewing of digital images by the end users. The utility tools developed by the CIL can be summarized as:

- Image Viewer for Digital Images
    - View by Thumbnail
    - Slide Show
    - View All
    - E-Greeting
- Devnarayan Zoom Tool
- Cognitive Mapping Tool
    - Cognitive Map of Muktesvara Temple
        - Decorative Motifs
        - Iconography
        - Architecture
- Hot Spot Viewer (a tool for 3D-viewing of photographs)

CIL has technical expertise to train digitization staff of partner institutions and similar projects on handling of digitization hardware and software tools.

4.2.2 Digitization Projects Implemented by IGNCA

IGNCA is the central hub of digitization and interactive multimedia documentation activities for Indian cultural heritage collections. Since 1994 IGNCA is engaged in cultural informatics with the capacity development project support from the United Nations Development Programme (UNDP). When the concept of cultural informatics shaped up in India, union Ministry of Communications and Information Technology (MCIT) in association with Ministry of Culture made provision of planned projects to capture rare documentary heritage collections in digital media. IGNCA was the obvious choice as an implementation agency for cultural informatics projects in partnership with central government-supported cultural institutions such as National Museum, National Museum Institute, Khuda Bux Oriental Public Library, etc. IGNCA has already implemented two MCIT-funded projects namely, Kalasampada: Digital Library - Resources of Indian Cultural Heritage (DL-RICH) and Cultural Heritage Digital Library in Hindi (CHDLH). Another MCIT-supported project, being planned and implemented at IGNCA, is National Databank on Indian Art and Culture (NDIAC).

Figure 4.5 shows a schematic structure of digitization and digital library projects in IGNCA where project funding body MCIT occupies top layer, followed by implementing agencies and implementation division. Figure 4.6 indicates three projects – one ongoing and two completed digitization projects under the auspices of MCIT.

Three major projects carried out in IGNCA created a corpus of about a million of digitized folios from documentary heritage collections. Table 4.1 provides a detailed account of scope and coverage of digitization and digital library projects in IGNCA.

In Kalasampada – DL-RICH project, digitized collections include: over 14,000 rolls of digitized microfilms of manuscripts; over one lakh (0.1 million) digitized slides, covering images of rare paintings, heritage sites, antiquities and artifacts; over 4000 rare photographs, having special collections of legendary photo-artists Rajah Deen Dayal and Sambhu Nath Saha; over 13,000 text pages of Sanskrit Text Repository (Gaudiya Grantha Mandira) in PDF format; over 1000 hours of audio and video recordings; selected IGNCA publications; back volumes of IGNCA newsletter *Vihangama* and journal *Kalakalpa*; over fifty virtual walkthroughs of selected monuments and heritage sites.

In Cultural Heritage Digital Library in Hindi (CHDLH), digitized collections include: over 10,000 textual pages in Hindi language, over 5000 images of antiquities, and over hundred hours of audio and video recordings.

Archaeological Survey of India (ASI) is a major stakeholder in ongoing project at IGNCA titled 'National Databank on Indian Art and Culture'. In this project proposed digitized collections will include: over 25,000 digitized rare books on Indian art and culture from ASI collection; over one lakh (0.1 million) digitized images of antiquities; over 1000 hours of audio and video recordings; all back volumes of ASI journal *Indian Archaeology–A Review*; and virtual walkthroughs of selected monuments and heritage sites in India.

Digitization projects, sponsored by MCIT, are described in details in later part of this chapter. In addition to the MCIT–supported digitization projects, IGNCA also undertakes turnkey digitization projects of different memory institutions of the country and abroad.

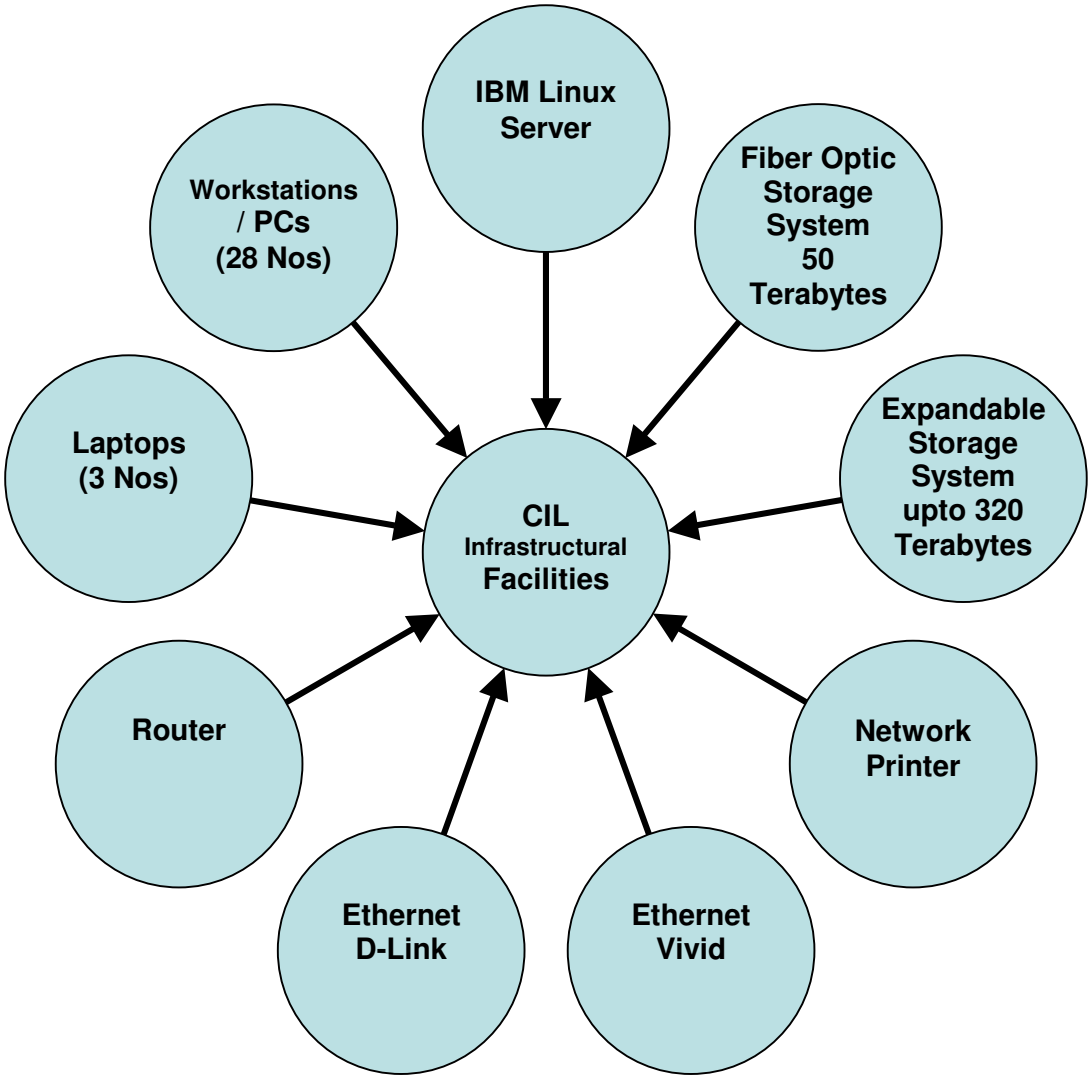


Fig. 4.2: ICT Infrastructural Facilities Available at CIL

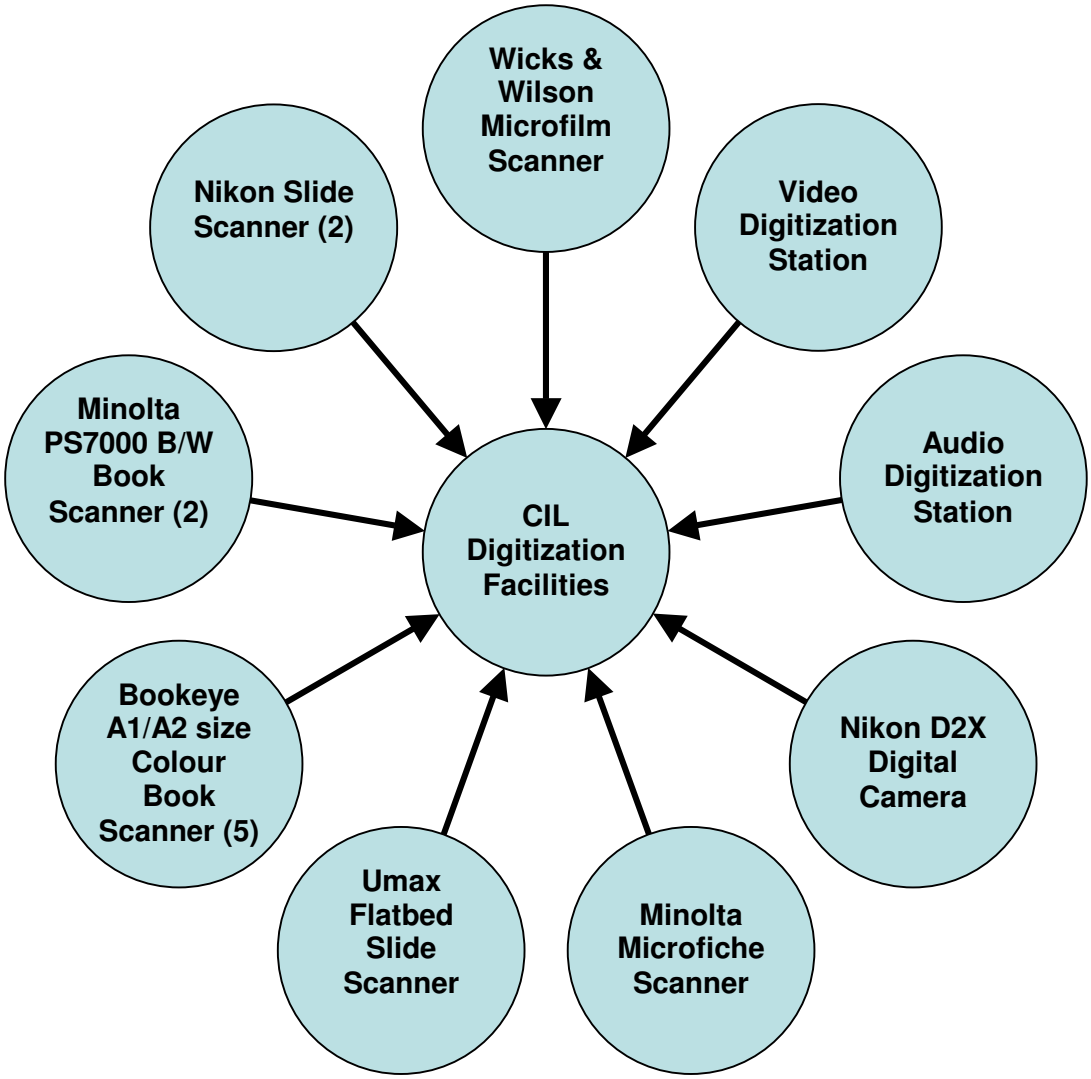


Fig. 4.3: Digitization Equipment and Facilities Available at CIL

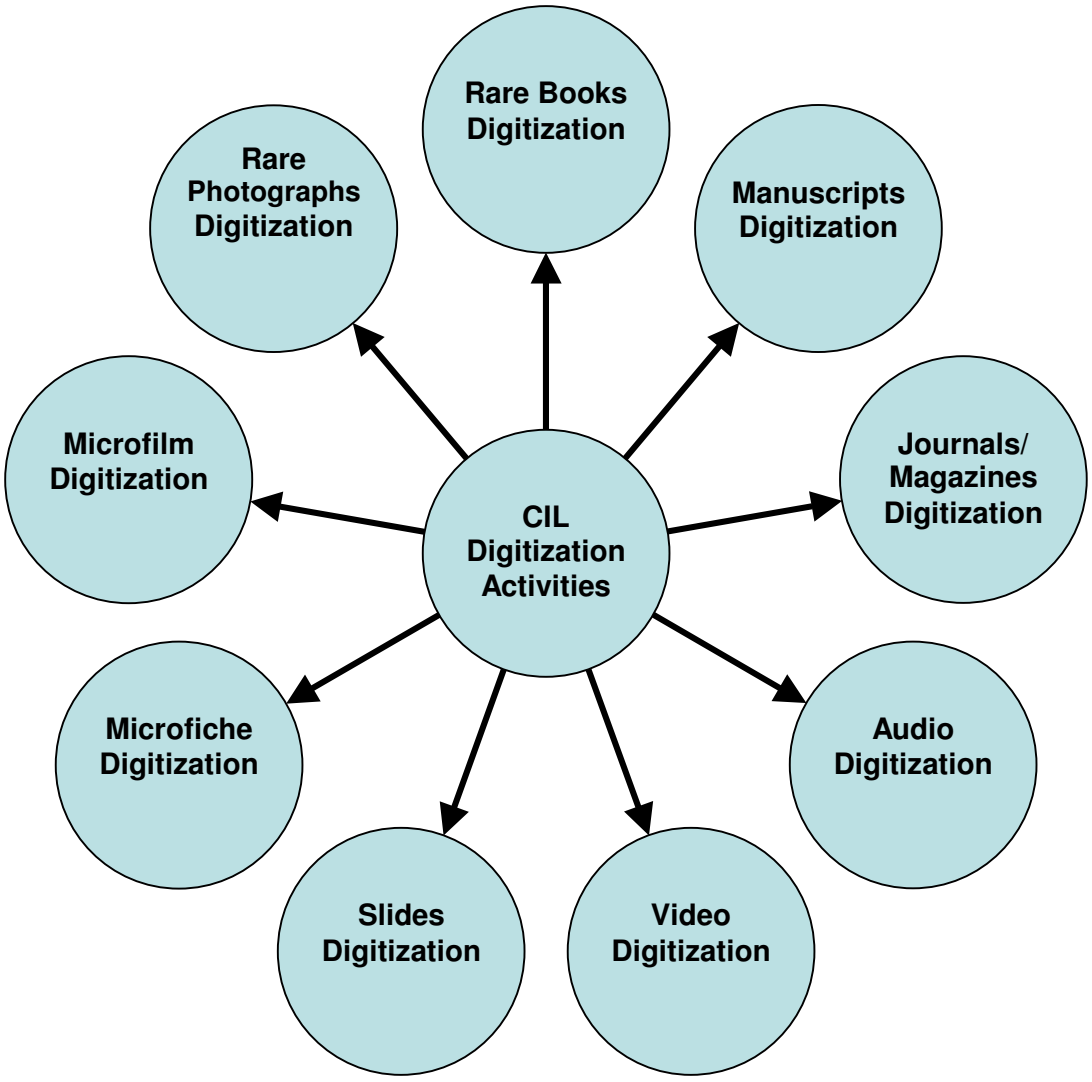


Fig. 4.4: CIL Digitization Activities

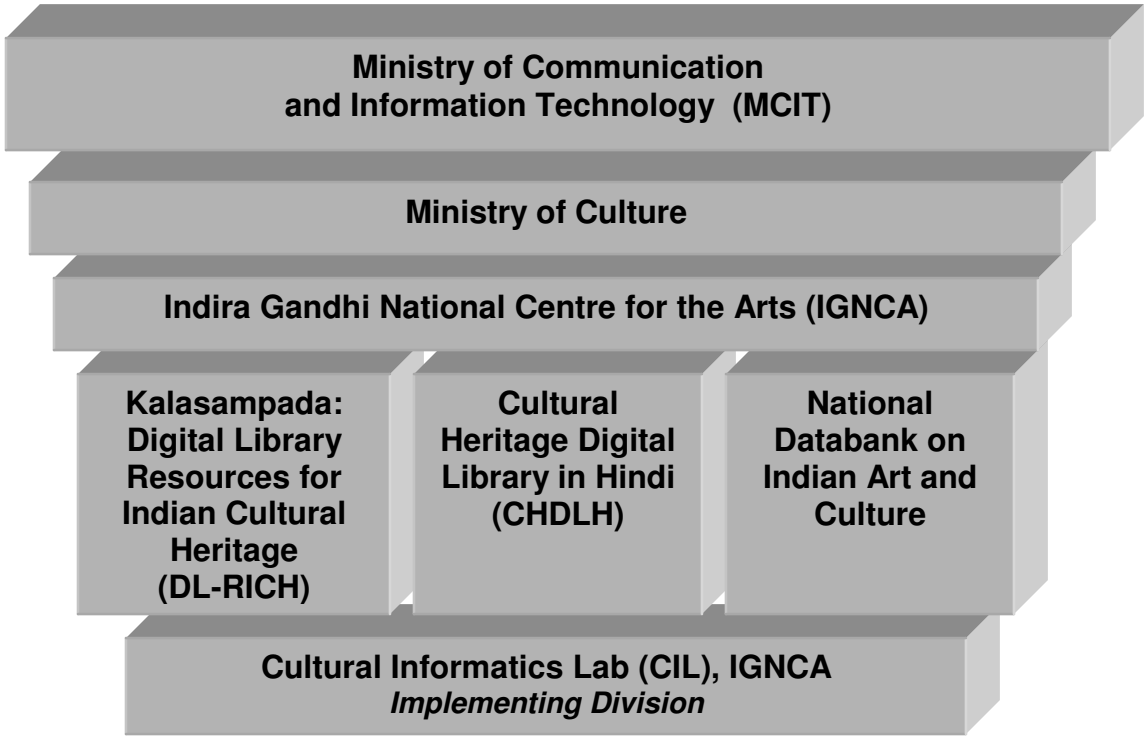


Fig. 4.5: Schematic Structure of Digital Library Projects in IGNCA

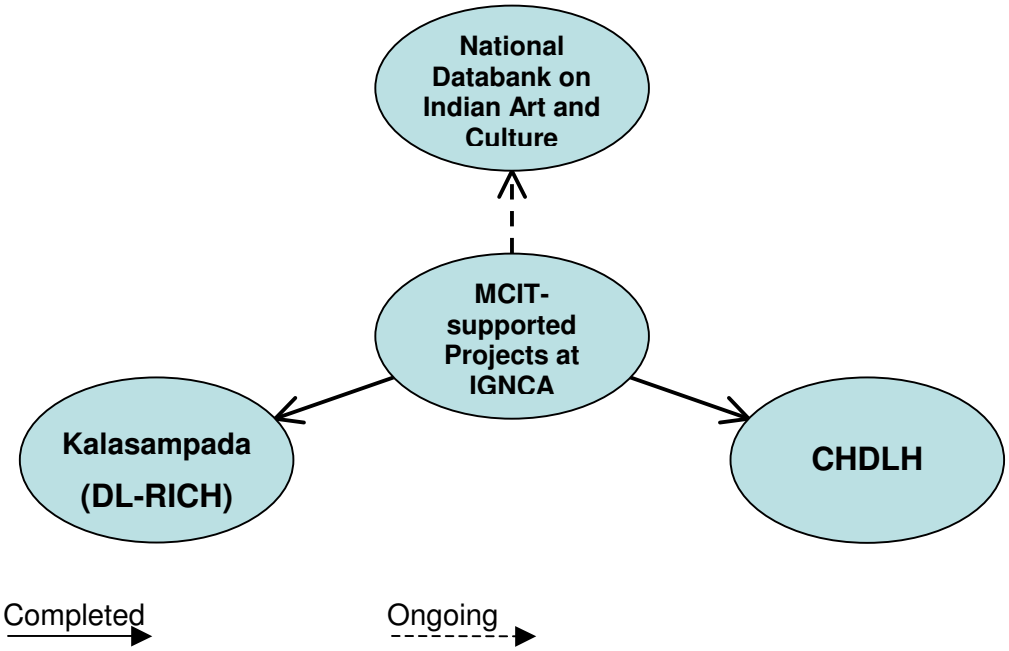


Fig. 4.6: MCIT-supported Projects at IGNCA

Table 4.1: Coverage of Digitization and Digital Library Projects in IGNCA

Name of Project	Kalasampada: Digital Library Resources for Indian Cultural Heritage (DL-RICH)	Cultural Heritage Digital Library in Hindi (CHDLH)	National Databank on Indian Art and Culture (NDIAC)
Coverage	• 14,000+ rolls of digitized microfilms of manuscripts • 100,000+ digitized slides, covering images of rare paintings, heritage sites, antiquities and artifacts • 4000+ rare photographs • 13,000+ pages text of Sanskrit Text Repository (Gaudiya Grantha Mandira) in PDF format • 1000+ hours of audio and video recordings. • Selected IGNCA publications; back volumes of IGNCA newsletter and journal • 50+ Virtual Walkthroughs of selected monuments and heritage sites.	• 10,000+ textual pages • 5000+ Images of antiquities • 100+ hours of audio and video recordings.	• 25,000+ digitized rare books on Indian art and culture • 100,000+ images of antiquities • 1000+ hours of audio and video recordings • Back Volumes of ASI Journal *Indian Archaeology – A Review* • Virtual Walkthroughs of selected monuments and heritage sites.
Status	Completed. New collections are added in regular interval.	Completed	Ongoing

4.2.3 Turnkey Digitization Projects at IGNCA

IGNCA has excellent digitization facilities that attract external institutions to avail digitization solutions and digitization services from IGNCA. It provides total digitization solutions to member institutions, starting from scanning of physical resources, to metadata creation, and archiving either in local server or Web-server. Figure 4.7 shows the names of clienteles availing digitization services.

The National Library of Mongolia in Ulaanbaatar is one of the few international clients of IGNCA that availed digitization services of IGNCA in collaboration with Indian Council for the Cultural Relations (ICCR). In this collaboration, Kanjuur and Danjuur manuscripts were digitized and made available at the premises of National Library of Mongolia for local access.

National Museum in New Delhi is another prestigious client of IGNCA. In an ongoing digitization programme, IGNCA has digitized over 2700 manuscripts of the National Museum collection, covering more than nine lakh (0.9 million) folios till 2007.

Three oriental institutions and manuscript libraries in Srinagar, namely Oriental Research Library, Shri Pratap Singh Library and Allama Iqbal Library, have collaborated with IGNCA under its manuscripts acquisition programme for digitizing all the manuscripts available with them. IGNCA digitized 10,581 Persian and Sanskrit manuscripts containing over 21 lakh (2.1 million) folios. The collection includes rare manuscripts on Yogavashistha, Mahabharata, Saivism, and Ayurveda written in Sanskrit (Sharada and Devnagari script). Later, another institution from Srinagar - Cultural Academy joined hand with IGNCA for digitizing its collection of antiquities.

Visva-Bharati in Santiniketan maintains a valuable collection of original manuscripts of India's first Nobel laureate Gurudev Rabindra Nath Tagore along with his personal letters, paintings and photographs. Visva-Bharati collaborated with IGNCA to avail its digitization services. IGNCA digitized over 76,000 folios of text covering 870 Tagore manuscripts and other antiquities related to Rabindra Nath Tagore.

IGNCA collaborated with Allahabad Museum in Uttar Pradesh for digitization of antiquity collections of Allahabad Museum containing over 6000 rare manuscripts, over 1500 sculptures and terracotta, and over 1000 paintings and photographs.

Recently, IGNCA extended its digitization solutions to Asiatic Society in Kolkata, Department of Historical and Antiquarian Studies in Guwahati, and Sri Chandra Shekharendra Saraswathi Viswa Maha Vidyalaya in Kanchipuram for digitizing their antiquity collections including manuscripts.

Table 4.2 provides an indicative list of turnkey projects undertaken by CIL. All of these turnkey projects were beyond the scopes of MCIT-supported digitization projects. This Table also indicates that IGNCA is very keen to take up consultancy projects related to digitization of documentary heritages utilizing its infrastructure and resources.

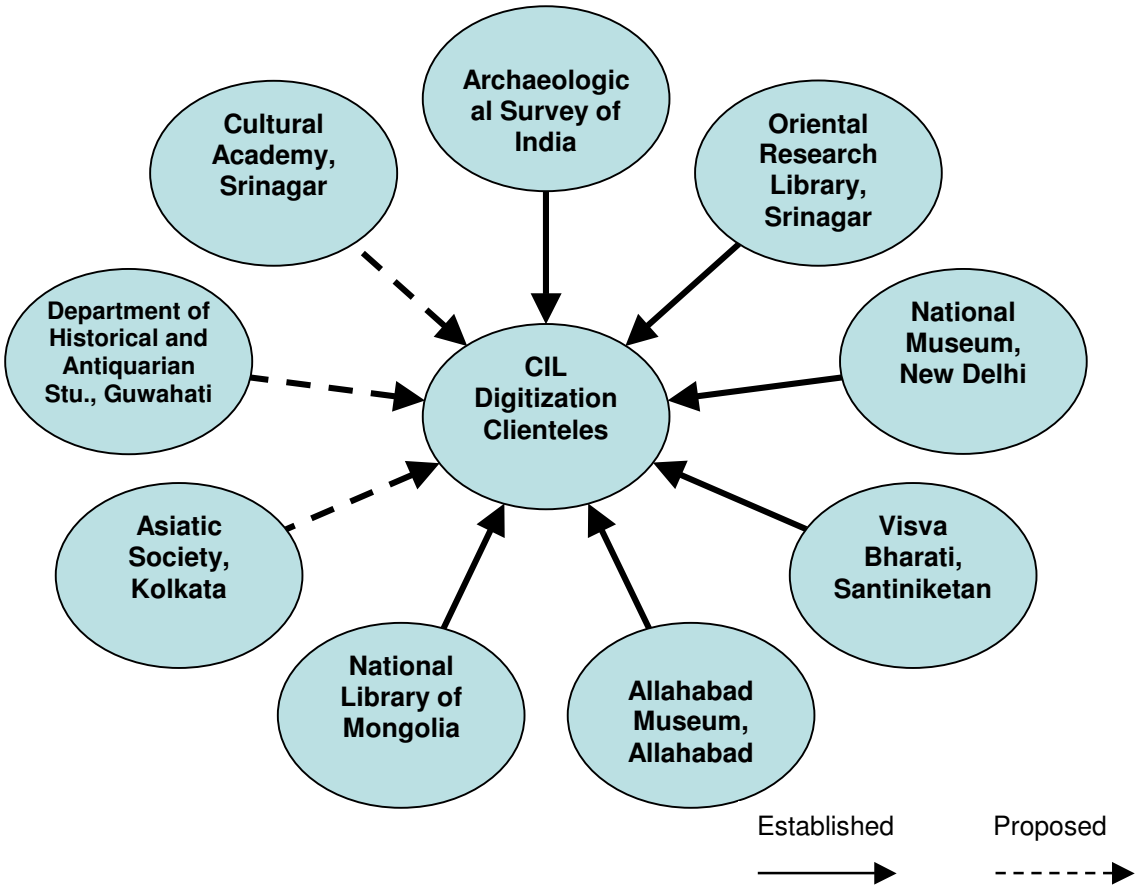


Fig. 4.7: CIL Digitization Clienteles

Table 4.2: Turnkey Digitization Projects undertaken by CIL

Name of Institution	Type of Materials	Coverage/ Status
National Museum, New Delhi	Digitization of manuscripts	2700+ manuscripts covering over 0.9 million folios.
Oriental Research Library, Srinagar	Digitization of Sanskrit and Persian manuscripts on Yogavashistha, Mahabharata, Saivism, Ayurveda, etc. in Sanskrit (Sharada and Devnagari scripts)	10,581 manuscripts covering over 2.1 million folios.
Shri Pratap Singh Library, Srinagar		
Allama Iqbal Library, Srinagar		
Visva-Bharati, Santiniketan	Digitization of antiquities, covering rare manuscripts of Rabindranath Tagore, his paintings, photographs and letters	76,000+ folios of text; 870 Tagore manuscripts
Allahabad Museum, Allahabad	Digitization of manuscripts and other antiquities	Over 6000 manuscripts, over 1500 sculptures and terracotta, over 1000 paintings and photographs.
Archaeological Survey of India (ASI), New Delhi	Digitization of rare books	Target: over 25,000 rare books
National Library of Mongolia, Ulaanbaatar	Digitization of Ganjuur and Danjuur manuscripts	Completed
The Asiatic Society, Kolkata	Digitization of manuscripts and other antiquities	Ongoing
Department of Historical and Antiquarian Studies, Guwahati, Assam	Digitization of manuscripts and other antiquities	Ongoing
Cultural Academy, Srinagar	Digitization of manuscripts and other antiquities	Ongoing
Sri Chandra Shekharendra Saraswathi Viswa Maha Vidyalaya, Kanchipuram, Tamil Nadu	Digitization of manuscripts and other antiquities	Ongoing

4.2.4 Collaboration Pattern in Digitization Projects at CIL

Over the time, digitization projects at CIL help in building up successful multi-stake partnership, where different kinds of agencies joined hands to develop synergies in developing large digital libraries of the cultural resources on Indian art, culture and heritage. Ministry of Communication and Information Technology (MCIT) is the supporting agency for three national projects being planned and implemented by IGNCA. CIL also has established technical collaboration with two mission mode flagship programmes of MCIT, named as 'Technology Development for Indian Languages Programme' (TDIL) and 'Content Development and IT Localization Network Programme' (CoIL-Net). Some of digitization equipment and tools now available with IGNCA were procured utilizing MCIT grants.

As parent body of IGNCA, Ministry of Culture is very supportive of the activities of CIL in terms of providing matching grants for infrastructural and capacity development in this unit. The UNDP (United Nations Development Programme) and UNESCO (United Nations Educational Scientific and Cultural Organization) also have established linkage with CIL through technical collaboration. Many organizations under Ministry of Culture have received technical support and technical guidance from IGNCA for planning and implementing their digitization programmes. This includes Archaeological Survey of India (ASI), National Museum, National Museum Institute, Asiatic Society in Kolkata, etc. There are numerous partner institutions which have collaborated with IGNCA for digitization of their documentary heritage collections. The partners were drawn from academic institutions, manuscript libraries, oriental institutions, religious institutions, government agencies, museums and archives. Figure 4.8 depicts collaboration pattern of IGNCA with different kinds of organizations and cultural institutions. IGNCA also has multi-layered collaboration with an array of international memory institutions, which lend their oriental materials for collection building of digital libraries at IGNCA. An indicative list of such international institutions is mentioned below:

- American Committee for South Asian Art (ACSAA), USA
- Los Angles CCU Museum of Art, California, USA
- The Cleveland Museum, USA
- The Skanhe Collection, University of Michigan, USA
- Nelson Aktins Museum, Kansas City, USA
- Kimball Museum of Art, Forwarth, Texas, USA
- The Asian Arts Museum of San Francisco, California, USA
- The India House Library, London, UK
- The British Library, London, UK

Digitization projects at IGNCA successfully adopted centralized digitization and dissemination model for digitizing a mass volume of cultural heritage resources. This project also helped in formation of a specialized segment of IT applications in India – called Cultural Informatics – to deal with digitization and interactive multimedia documentation services.

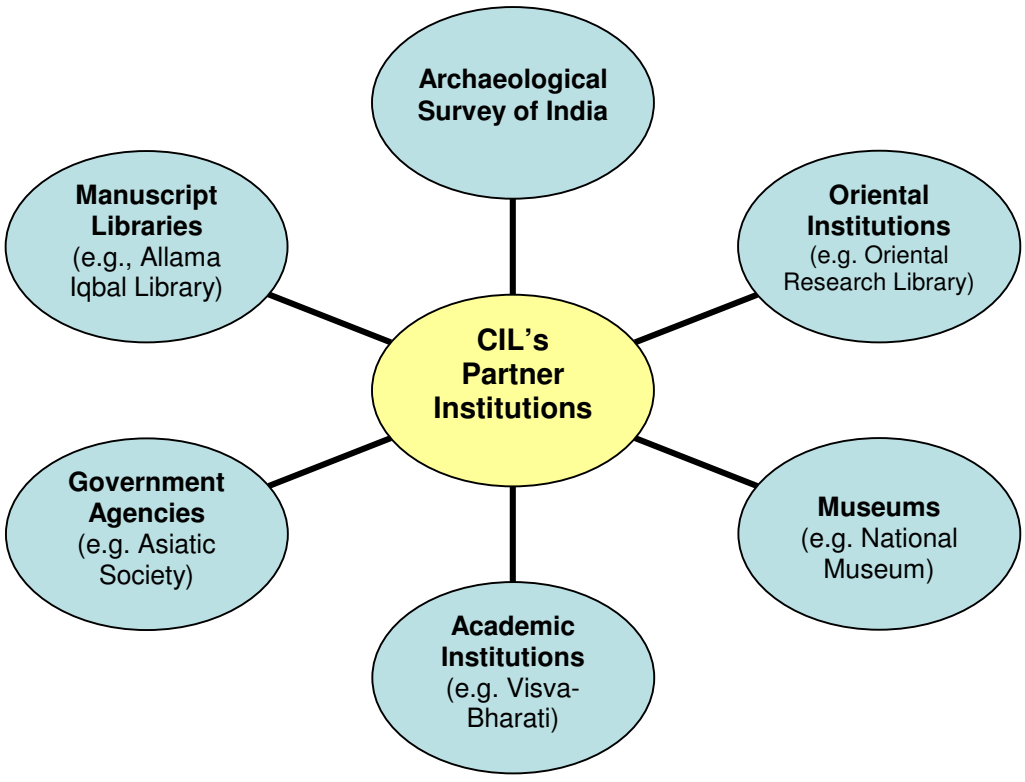


Fig. 4.8: Collaboration Pattern at IGNCA for Digitization Solutions

4.2.5 Digitization Standards as Followed in IGNCA

For any kind of digital object there are some proprietary formats as well as open formats. A proprietary format allows only a specific computer application or an application suite for interacting with the digital objects. On the other hand, an open format allows interacting with digital objects using any generic application meant for that kind of digital object.

IGNCA maintains consistency in its digitization projects in terms of conformity to international standards, particularly conformity to open formats for created digital objects. IGNCA usually follows stringent quality norms at the output level to make provision of reuse of created digital objects. Digitization standards facilitate cross-walk and interoperability across different applications and retrieval systems. Digitization standards, as followed at the IGNCA for various kinds of materials, conform to UNESCO Guidelines[1] published in March 2002 (UNESCO, 2002).

Sometimes, high quality digital objects are not suitable for Web-based online access since many Internet users in India do not have broadband connectivity. Thus, lower quality derivatives are used for the online access of the materials. Original digital images are primarily saved in TIFF (Tagged Image File Format) files – which are uncompressed storage mode for digital images. On the other hand, photo slides are digitized and saved in Photo CD format (PCD) – another uncompressed storage mode. Later digital images in TIFF or PCD formats are converted into compressed image formats such as JPEG format, suitable for Web-based applications. Digitization standards as followed in IGNCA for different kinds of materials are shown in Table 4.3.

Figure 4.9 shows an image of manuscript folio that was scanned and saved in TIFF format and later converted into JPEG format for making it Web accessible.

[1] Guidelines for Digitization Projects for Collections and Holdings in the Public Domain, Particularly Those Held by Libraries and Archives; jointly prepared by UNESCO, IFLA and International Council on Archives, March 2002. Online accessible at: http://www.ifla.org/VII/s19/pubs/digit-guide.pdf

Table 4.3: Digitization Standards as followed at the IGNCA for Different Types of Materials

Type of Material	Standards
Slides	Photo CD Format[1] (five resolutions normally and six resolutions in specific cases)·
Microforms (Microfiche and Microfilms)	300 dpi[2] TIFF[3]
Photographs	300 dpi TIFF (600 / 1200 dpi in special cases)
Books and printed materials	300 dpi TIFF
Audio	.WAV file format[4] with 44 Khz sampling rate[5]
Video	MPEG 1or MPEG 2[6]

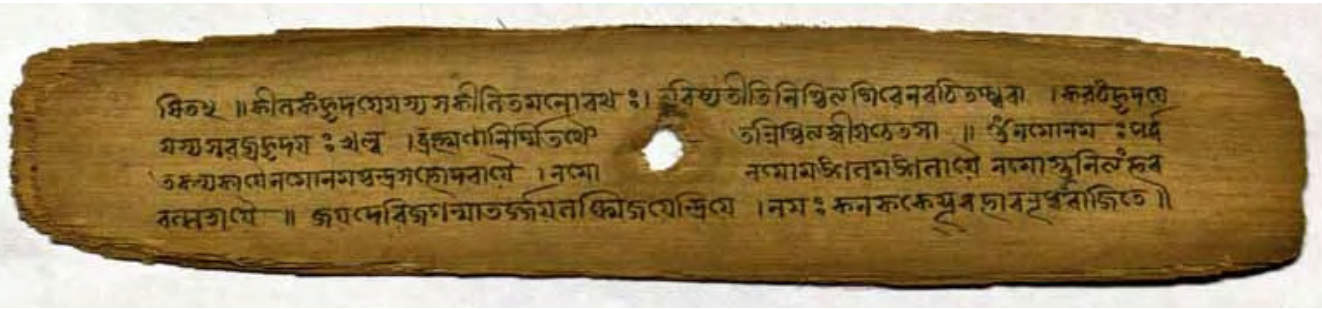


Fig. 4.9: Palm leaf Manuscript of Durga Saptashati, digitized by IGNCA

[1] Photo CD: a system designed by Kodak for digitizing and storing photos in a CD.

[2] DPI: Dots per Inch, a measure of printing resolution.

[3] TIFF: Tagged Image File Format – an uncompressed storage mode for digital images

[4] WAV: Waveform audio format, a Microsoft and IBM audio file format standard for storing an audio bitstream on PCs.

[5] Sample rate: number of samples of a sound that are taken per second to represent sound digitally. The more samples taken per second, the better the reproduction of the sound. The current sample rate for CD-quality audio is 44,100 samples per second, or 44.1 kHz.

[6] MPEG-1: Video compression standard used for creation of video CD.

MPEG-2: Video compression standard for creation of broadcast-quality television videos that can be distributed on DVD and similar disks.

4.3 Kalasampada: Digital Library - Resources of Indian Cultural Heritage (DL-RICH) [www.ignca.nic.in/dgt_0001.htm]

The Indira Gandhi National Centre for the Arts (IGNCA) established a digital library on indigenous cultural heritage, namely 'Kalasampada: Digital Library – Resource for Indian Cultural Heritage' (DL-RICH) with the support from India's Ministry of Communications and Information Technology (MCIT). Figure 4.10 provides a schematic structure of DL-RICH project, where MCIT occupies top layer being a funding partner of the project. This project was monitored by National Digital Libraries Cell on behalf of MCIT. Cultural Informatics Lab in IGNCA was the implementing partner of this project – being the central agency for content creation through digitization and content dissemination through Web-based portal and intranet-based digital library system.

DL-RICH aims at using multimedia computer technology integrating variety of cultural information, and helping users to interact and explore the subjects available in DL-RICH collections on a computer in a non-linear mode by a click of mouse. Figure 4.11 illustrates project aims of DL-RICH.

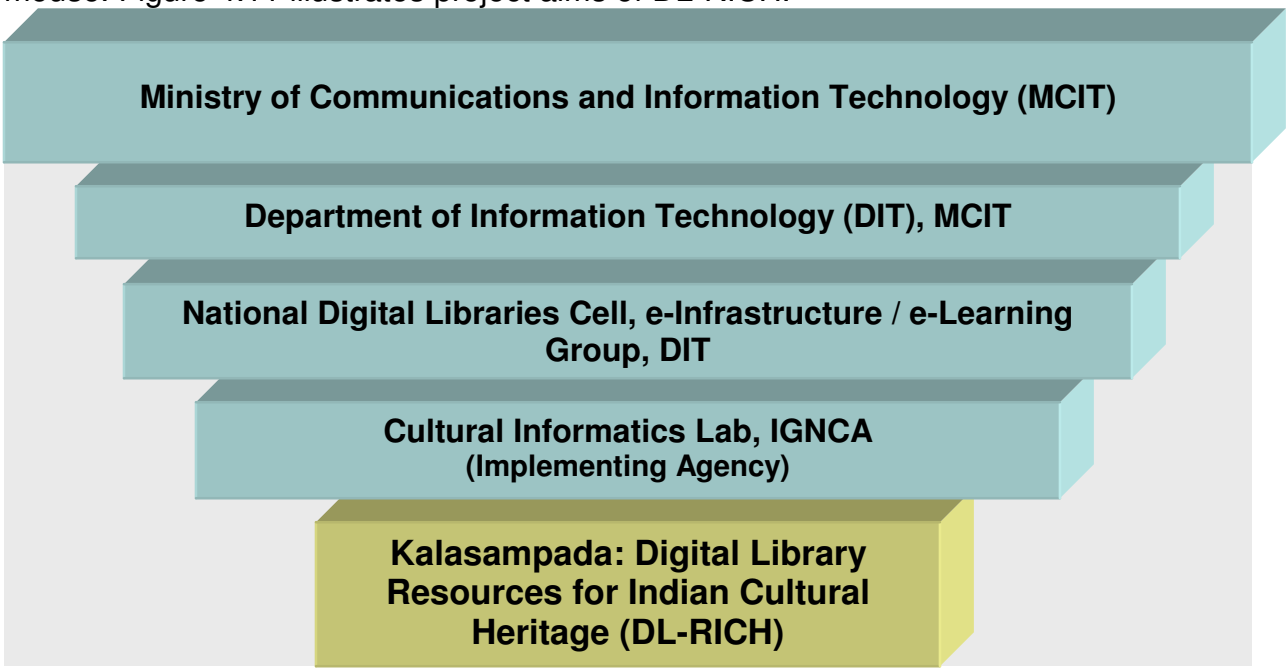


Fig. 4.10: Schematic Structure of DL-RICH Project

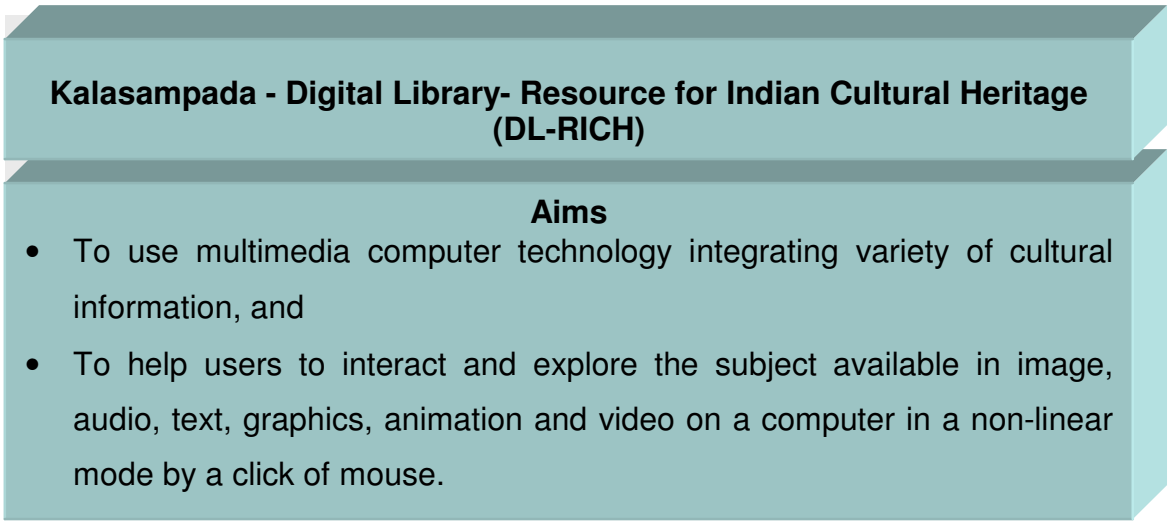


Fig. 4.11: Aims of DL-RICH Project

4.3.1 Digitized Collections in DL-RICH

Since its inception, DL-RICH is recognized as a reference databank for cultural heritage of India to encompass and preserve the distributed fragments of Indian arts, aesthetics and culture. This huge knowledge base helps the scholars to explore and visualize the information stored in multiple layers. DL-RICH provides online access to digital images of cultural heritage resources such as manuscripts, rare photographs, rare books, rare painting, sculptures, handicrafts, monuments, artifacts, festivals, as well as varieties of textual, graphical, audio-visual and multimedia resources. It hosts a Sanskrit text repository, called Gaudiya Grantha Mandira, covering more than 400 chapters from different oriental texts. The digital corpus in this knowledge base includes over five lakh (0.5 million) folios of manuscripts, slides, digitized photographs, IGNCA published books, periodicals *Kalakalp* and *Vihangama*, over 400 hours of audio and video, and approximately 50 walkthroughs. Table 4.4 differentiates the multimedia collections from textual collections available in DL-RICH portal. Some of the documents in the collections are not available to the open access domain due to copyright restrictions.

DL-RICH can be navigated through the hyper-linked indexes in the main page, thematic categories, and geographic categories as indicated in Table 4.5. The index to thematic categories gives further links to documents pertaining to literature, personalities, catalogues, performances, studies, events, heritage, collections, religious categories, etc. The index to geographic categories gives further links to states and union territories of India, and other countries. Although DL-RICH portal does not provide a comprehensive search facility, user-friendly interfaces in this portal make navigation convenient to the end users.

4.3.2 Intranet-based Collections

DL-RICH is a huge knowledge base, produced in partnership with member institutions having original source materials. As IGNCA does not possess ownership of most of the original source materials – it has limitation in providing open access to all digitized materials particularly through Web portal. Thus, IGNCA developed an intranet-based digital library system that provides in-house access to all digital resources emanated from this project. Intranet-based retrieval application aims at providing a user-friendly interface that helps the scholars to explore and visualize the information stored in multiple layers. Search facility is available both in English and Hindi (Devanagari). The user has the option to select the material of his interest either from a specific type of collection such as books, manuscripts, slides, audio and video or from the entire collection. Some of the reasons restricting IGNCA to disseminate full resources of DL-RICH through Web portal are:

(i) These materials are priced possession and covered under intellectual property rights such as copyright.
(ii) IGNCA does not have necessary approvals from the source institutions for providing open access to worldwide communities.

Thus, IGNCA provides partial access to DL-RICH resources from the IGNCA's official Web portal, where digital resources uploaded with necessary approvals. A majority of these materials are available within campus on intranet only as

indicated in Figure 4.12. An intranet-based retrieval application is used for this purpose having a user-friendly interface as shown in Figure 4.13.

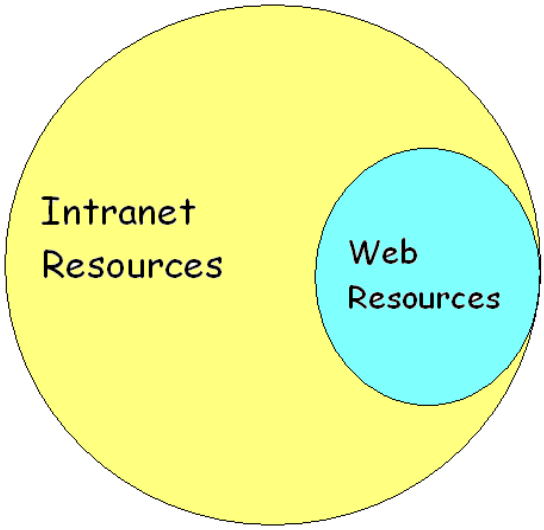


Fig. 4.12: DL-RICH Knowledge Base – A Mix of Closed Access and Open Access Resources

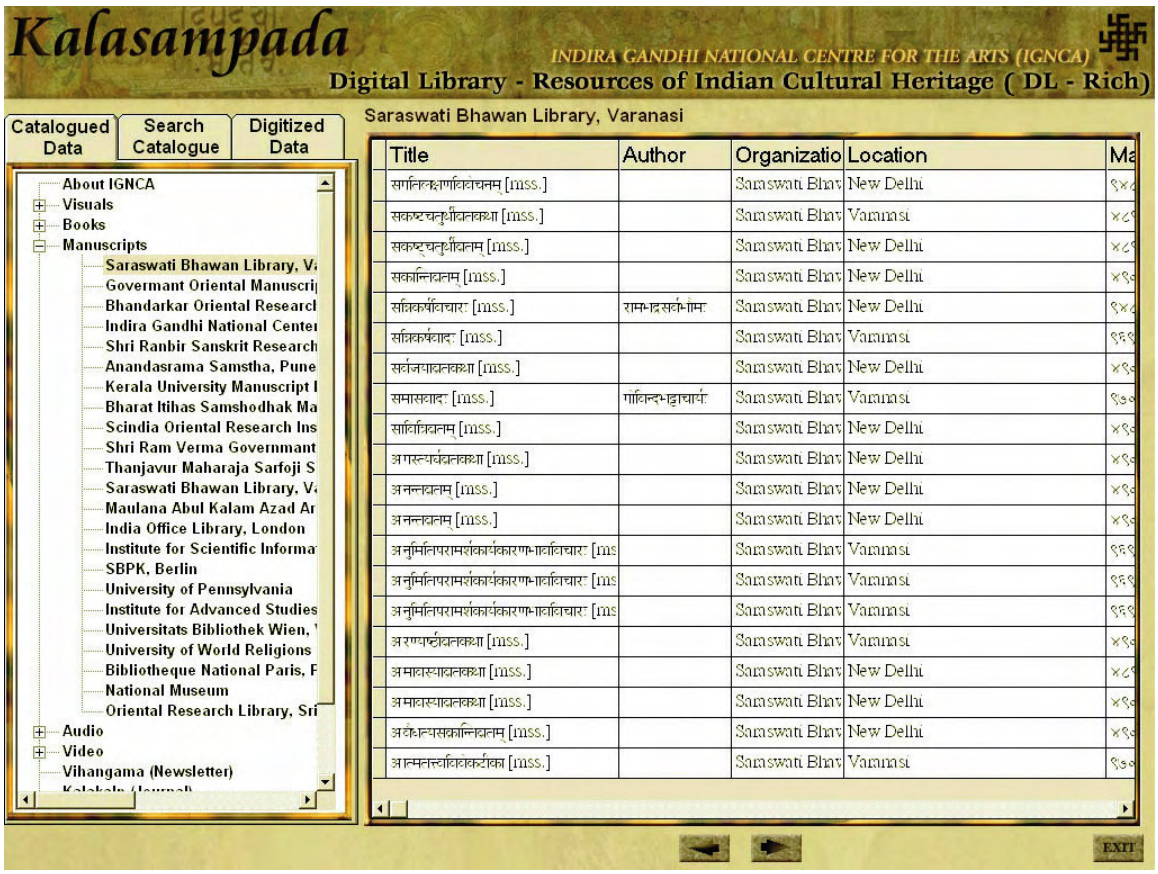


Fig. 4.13: Intranet-based Retrieval Application for DL-RICH Resources

Table 4.4: Different Kinds of Multimedia and Textual Collections in DL-RICH

Multimedia Collections	Textual Collections
Multimedia Documentation • The Illustrated Jataka & Other Stories of the Buddha • Art & Crafts of North-East Video Recordings • Gita Govinda • Brhadisvara Temple • Kutiyattam Audio Recordings • Indian Classical Music (Hindustani) • Muktesvara Temple CD-ROM : Devotional Music • Devnarayan Selected Contents of Multimedia CD-ROMs • Ajanta • Devadasi Murai • Muktesvara • Rockart • Cognitive Map of Muktesvara Temple Digital Images • Archaeological Sites • Exhibitions of Paintings • Photograph Collections • Slide Collections	Newsletter of IGNCA • Back volumes Journal of IGNCA • Back volumes Books Published by IGNCA • Electronic Books Sanskrit Text Repository • Gaudiya Grantha Mandira Research Reports Conference Proceedings Articles by In-house Scholars Papers and Essays Bibliographies Manuscript Catalogues • Catalogue of Microfilmed Manuscripts • Survey of Manuscripts in India Databases • Catalogue of Catalogues (CATCAT)

Table 4.5: Category-wise Collections in DL-RICH

Thematic Categories	Geographical Categories
Literature • Oral • Written Personalities • Dancers • Painters • Photographers • Scholars • Singers Performances • Dance • Music Studies • Life Style Events • Exhibitions Heritage • Monuments Collections • Coins • Cultural Archive • Decorative Arts • Masks • Paintings • Photographs • Puppets • Rare Books • Sculptures • Slides Religious Texts • Buddhism • Hinduism • Islamic • Jainism • Shaivism • Vaishnavism • Vedic tradition	States of India • Andhra Pradesh • Arunachal Pradesh • Assam • Bihar • Delhi • Gujarat • Haryana • Himachal Pradesh • Jammu & Kashmir • Karnataka • Kerala • Madhya Pradesh • Manipur • Meghalaya • Mizoram • Nagaland • Orissa • Punjab • Rajasthan • Tamil Nadu • Tripura • Uttar Pradesh • West Bengal Union Territories • Andaman and Nicobar Islands • Lakshadweep Other Countries • Australia • Bangladesh • China • Indonesia • Japan • Korea • Mongolia • Nepal • Sri Lanka • Thailand • Turkey • Vietnam

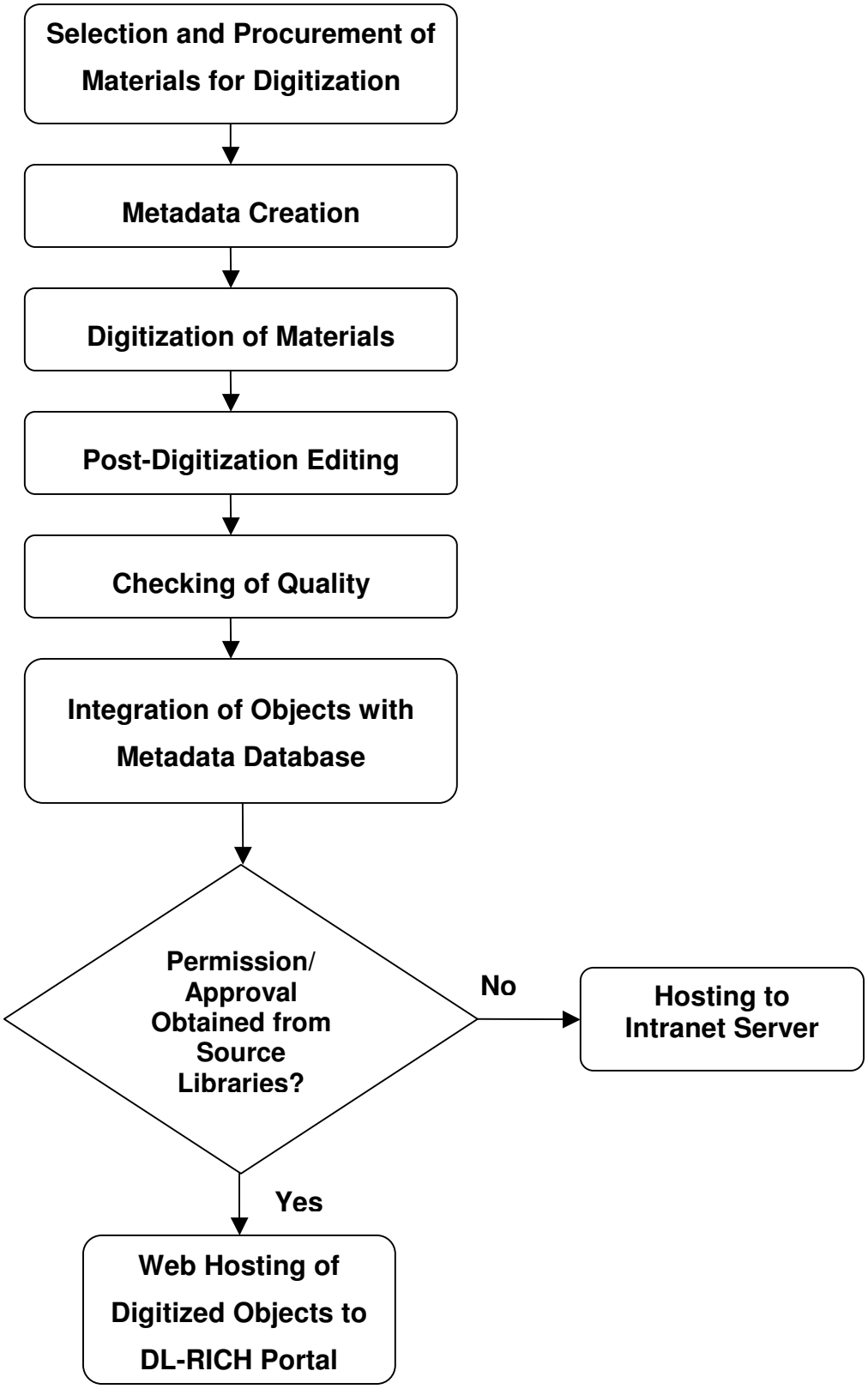


Fig. 4.14: Digitization Workflow in DL-RICH Project

4.3.3 Workflow in Creation of DL-RICH Resources

DL-RICH was the foremost cultural informatics project supported by MCIT and implemented by Cultural Informatics Lab in IGNCA. As a starting project of MCIT, it adopted international standards, norms and processes based on international guidelines available at that time in 2001. DL-RICH project has well-defined process workflow, where process workflow is divided into three major process elements, viz., digitization, post-digitization editing and integration. Each process element also has a number of pre-defined activities. Figure 4.14 depicts the DL-RICH process workflow, where input is a physical document and outcome is a digital document with essential metadata information. In between the input and outcome of the digitization process, rigorous procedures of standard compliance and quality assurance are applied to make the digital surrogates highly compatible for providing digital library service. Integration of digital surrogates and their metadata element sets with retrieval application is an essential process followed by compliance testing on effectiveness of integration.

The source institution usually signs memorandum of understanding (MoU), a legal agreement, with IGNCA to establish partnership between them for this project. This agreement is established to safeguard the joint ownership of digitized documents. Materials selected for digitization have much importance in the context of Indian cultural heritage and arts. At the end of digitization process of selected works, participating institution can opt for either open access through Web portal or limited access at the respective source institution in addition to IGNCA campus through intranet.

DL-RICH is one of the highly visible digital library portals in South Asia. DL-RICH received the 'Golden Icon: Award for Exemplary Implementation for e-Governance Initiatives' under the category, Best Documented Knowledge and case study, given by India's Department of Administrative Reforms and Public Grievances in 2005.

4.4 Cultural Heritage Digital Library in Hindi (CHDLH)

IGNCA has established another digital library on indigenous cultural heritage – the Cultural Heritage Digital Library in Hindi (CHDLH) – having local language content creation facility and delivery mechanism. This digital library was developed under the auspices of the Content Development and IT Localization Network (CoIL-Net) Programme of Ministry of Communication and Information Technology (MCIT) with special focus on the Hindi-speaking region in India (IGNCA, 2003, IGNCA, 2007).

4.4.1 CHDLH in National Context

Government of India launched a mission mode programme called the Technology Development for Indian Languages (TDIL) programme in early 1990s. This programme has overarching aim to integrate a large multilingual multicultural society of India in a common information sharing digital platform through the implementation of local language computing, IT localization and IT contextualization. Figure 4.16 illustrates the aim and objectives of TDIL Programme. This programme has contributed towards narrowing down communication barriers and bridging digital divide across the country through a number of intervention initiatives. This programme helped in development of many socially-useful systems and services such as (i) Indian Language Processing (ILP) Resources, (ii) Indian Language Processing (ILP) Tools, (iii) Translation Support Systems, and (iv) Human-Machine Interface Systems. Some of the end products emanated from this programme are multilingual dictionaries, multilingual thesauri, educational software, encyclopedia, creative writing system, translation support systems, OCR, text-to-speech (TTS) system, speech recognition system, pocket translator, personal digital assistants, reading machine for blinds & deaf and Indian language portals.

Although, Indian language applications and services were made available in late 1990s and later period through this programme, there was a scarcity of contents in Indian languages related and relevant to the grassroots communities. TDIL Programme took a focused initiative to solve this kind of problem. This initiative is

called 'Content Development and IT Localisation Network' (CoIL-NET). This initiative is formulated encompassing a vision for all pervasive socioeconomic developments by proliferating use of language specific IT-based contents, solutions and applications. This network was established for bringing the benefits of IT revolution to the common citizen and more specifically bridging the existing digital divide in the Hindi-speaking states of India. Figure 4.17 shows the aims and objectives of CoIL-Net Programme. Out of many objectives of CoIL-Net Programme, objectives emphasising need of digitization and digital library development in an Indian language are:

(i) To develop Cultural Heritage Digital Library in Hindi with special focus on the identified 7 states;

(ii) To promote content digitization for promoting access and sharing of public sector information in Hindi.

TDIL Programme is the primary collaborator of CHDLH initiative through its CoIL-Net Programme. Figure 4.15 provides a schematic structure of CHDLH, where TDIL Programme is placed in top layer in terms followed by CoIL-Net and implementing agency IGNCA.

CoIL-Net is a national network of eleven intervention initiatives for IT localization and content creation. Out of them four initiatives are national in scope and seven initiatives are state-specific – i.e., each initiative caters to a particular state out of seven Hindi-speaking target states. CHDLH is one of the successful national level intervention initiatives within CoIL-Net, where more contents are created for Web dissemination than any other network member. Figure 4.18 shows a list of national and state level intervention initiatives affiliated to CoIL-Net Programme, where CHDLH is placed very prominently not only in India but also to the Hindi-speaking diasporas abroad.

Figure 4.19 indicates that CHDLH aims at offering carefully selected thoughtfully compiled and contextually integrated multimedia content on cultural heritage, folk literature and lifestyles of Hindi-speaking region. This Figure also indicates that CHDLH is engaged in development of a reusable 'model design' and

'development process' for implementing user-friendly Web-enabled heritage library.

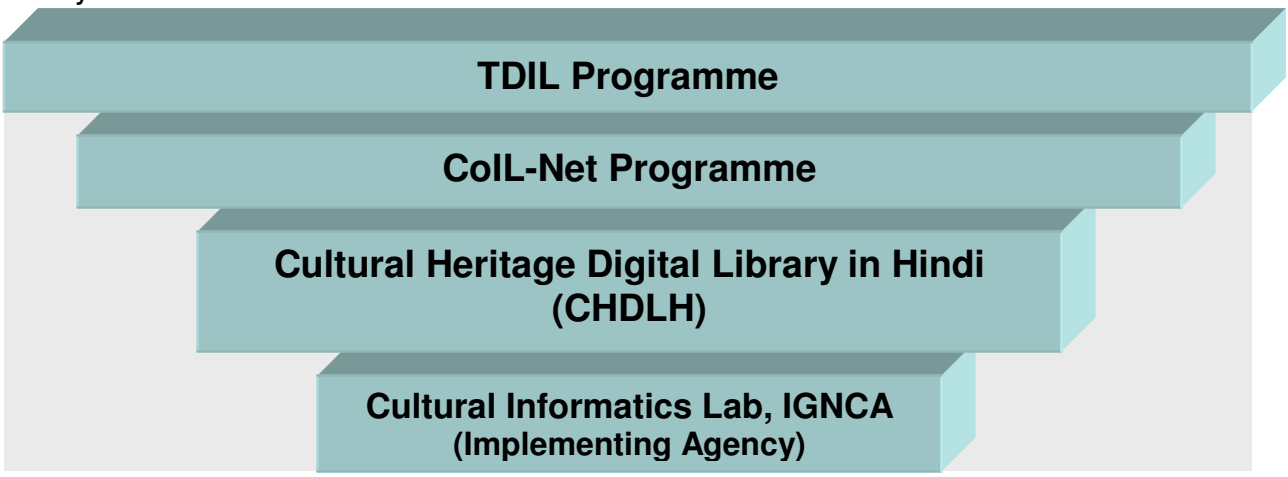
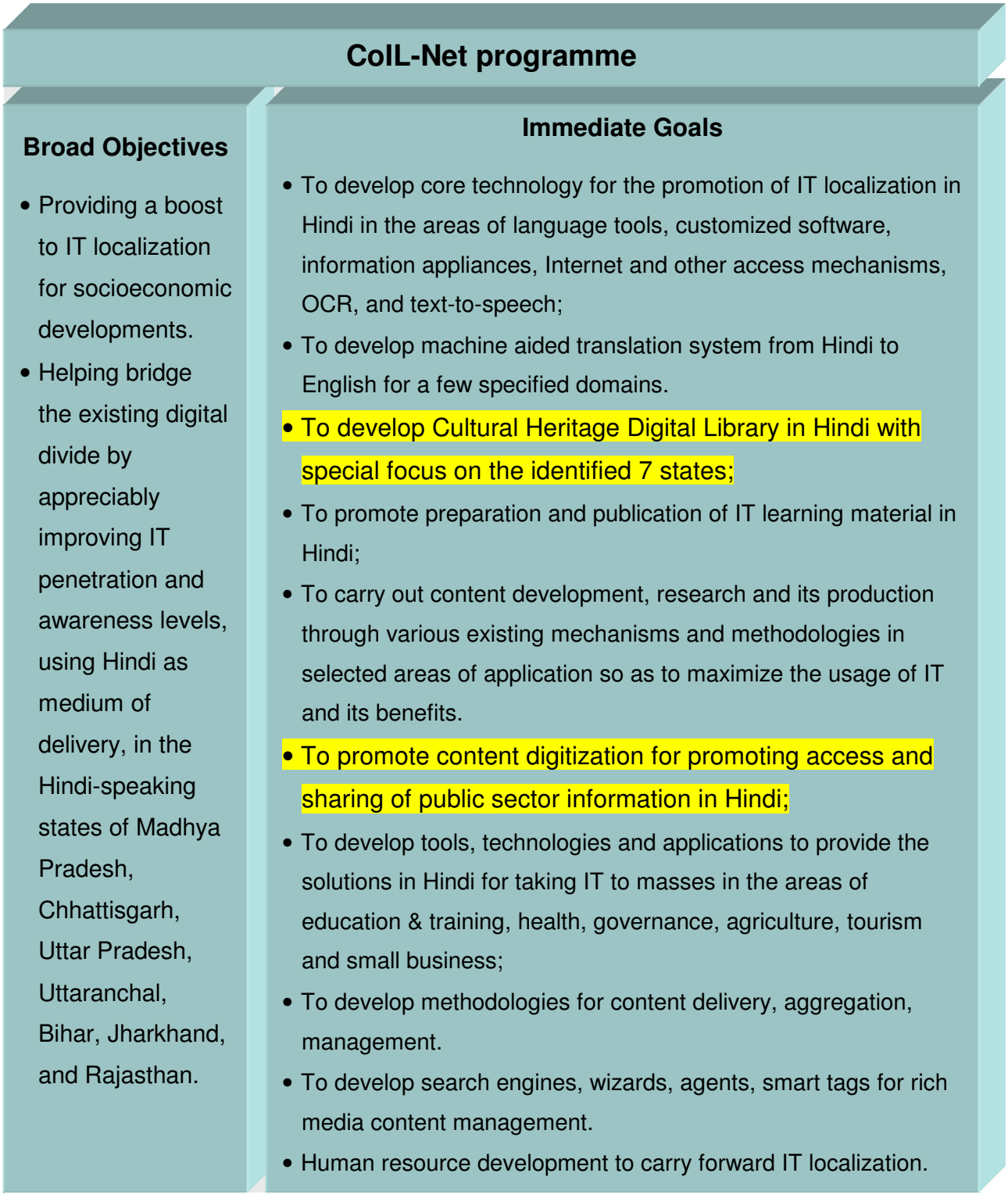


Fig. 4.15: Schematic Structure of CHDLH

TDIL Programme

Aim	Objectives
Promotion of IT tools for Indian Languages.	• To develop information processing tools to facilitate human machine interaction, in Indian languages and multi-lingual knowledge resources. • To promote the use of information processing tools for language studies and research. • To support R&D efforts in the area of information processing in Indian languages and to support research on knowledge tools: representation, integration, compression and learning methodologies. • To consolidate technologies thus developed for Indian languages and integrate these to develop innovative user products and services.

Source: *About TDIL Programme [http://tdil.mit.gov.in/tdilpindx.html]*



Source: *About CoIL-Net programme [http://tdil.mit.gov.in/coilnet.htm]*

Fig. 4.17: Broad Objectives and Immediate Goals of CoIL-Net Programme

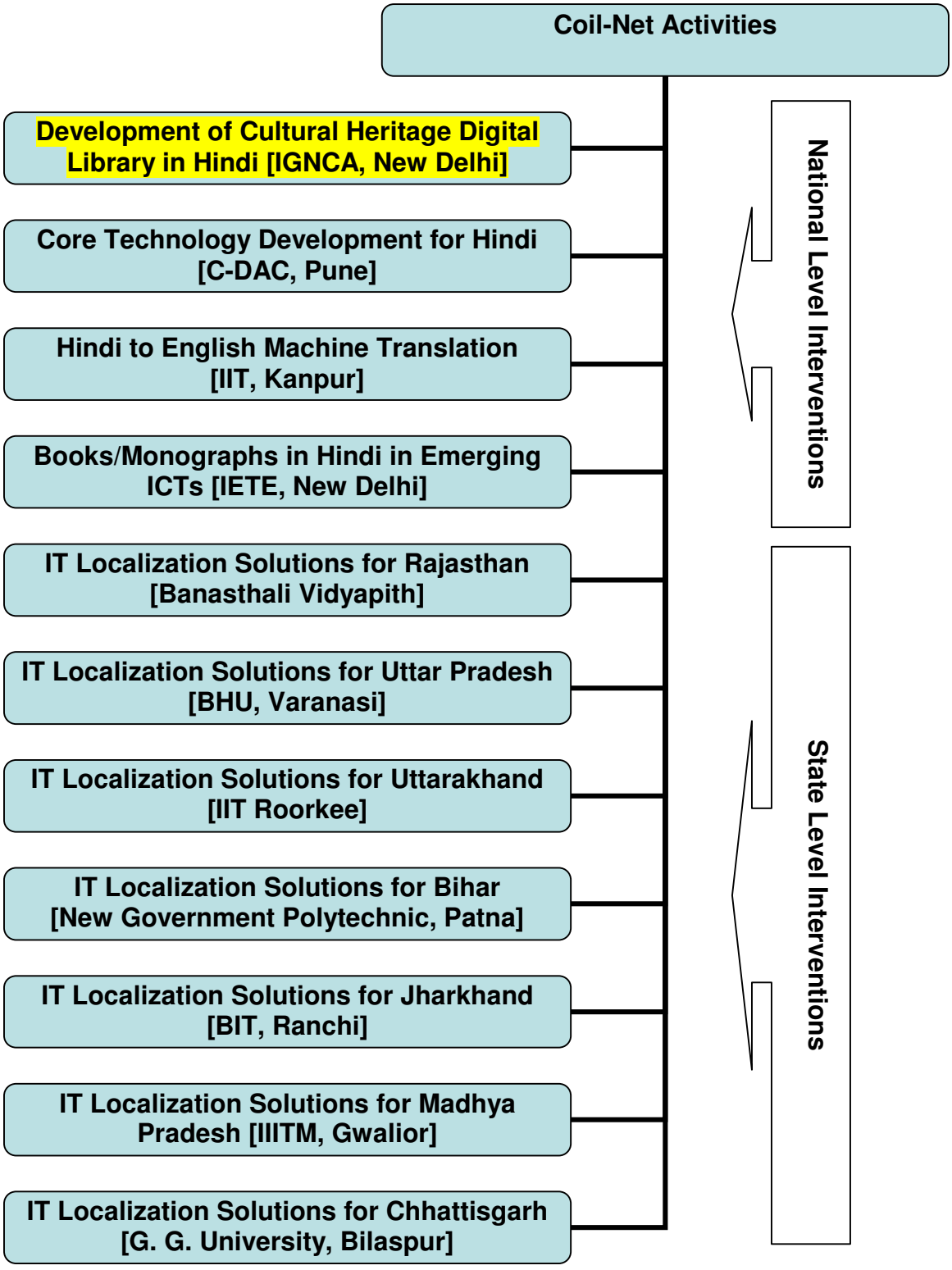
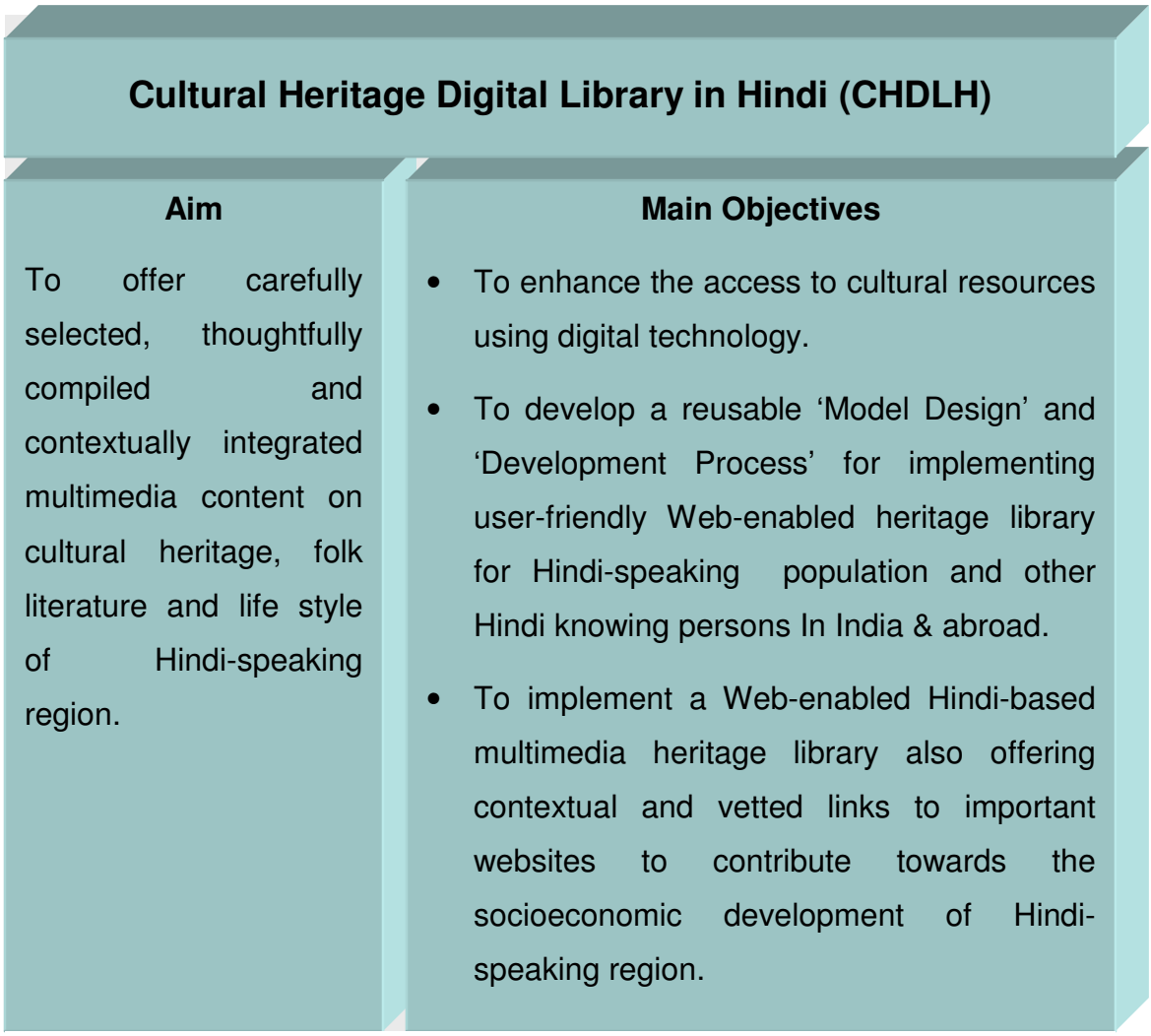


Fig. 4.18: National and State Level Interventions through CoIL-Net



Source: *About CoIL-Net <www.ignca.nic.in/mmd_body.htm#CoIL>*

Fig. 4.19: Aim and Main Objectives of CHDLH

4.4.2 Digitized Collections in CHDLH

CHDLH represents ethos of Indian heritage in general and Hindi-speaking target states in particular. In this project most of the contents has been compiled by the project investigators and scholars from the respective region covering the major topics. CHDLH disseminates information and traditional knowledge related to:

- Common heritage of the people;
- Poetic and literary heritage;
- Architectural heritage;

- Natural heritage; and
- Miscellaneous information related to other areas of arts, aesthetics and culture.

CHDLH provides access to varieties of textual, graphical and multimedia digital resources, almost the same as documents available in DL-RICH, but restricted to Hindi language. Contents of CHDLH are developed mainly in Hindi and displayed in Devanagari scripts using Unicode.

CHDLH collections also focus on cultural heritage, folk literature and lifestyle of each of the target states of Hindi-speaking region. These contents also portray cultural diversity and uniqueness of selected states of India. Figure 4.20 indicates the focused states of India that become both beneficiaries as well as content providers of this portal.

Similar to DL-RICH portal, CHDLH portal provides user-friendly browsing facilities through the hyper-linked index in the main page, index to thematic categories, and index to geographic categories. Similar to DL-RICH portal, CHDLH portal does not have search facility.

Table 4.6 indicates the available navigation options in the main page of this portal. This Table also indicates that different kinds of multimedia materials are integrated in this portal. Figure 4.21 demonstrates that CHDLH collections comprise a mix of multimedia and textual contents.

Table 4.6: Navigation Options Available with CHDLH Portal

Navigation Menu for CHDLH Portal	Navigation Menu in Hindi for CHDLH Portal
• Traditional Literature • Digital Images • Oral Epics • Folk Literature • Geographic Class • Documentation • Audio • Video • Books • Poets and Writers • Articles	

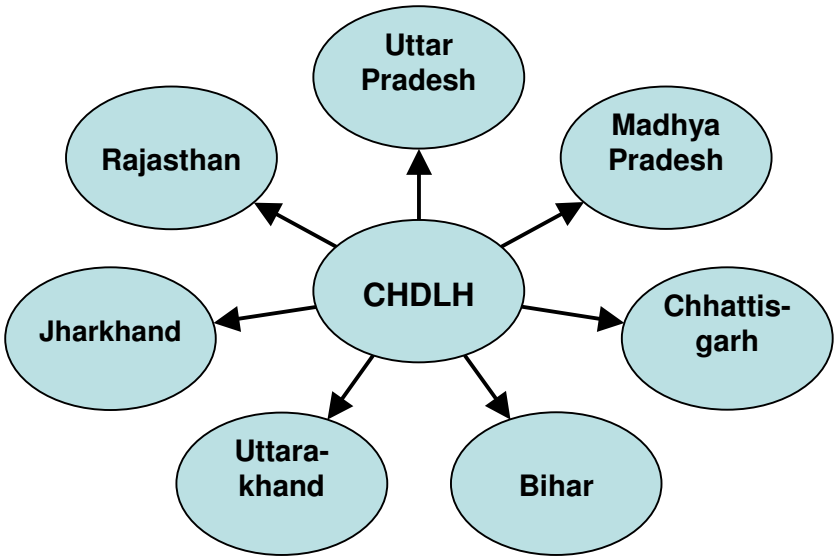


Fig. 4.20: Focused States of CHDLH

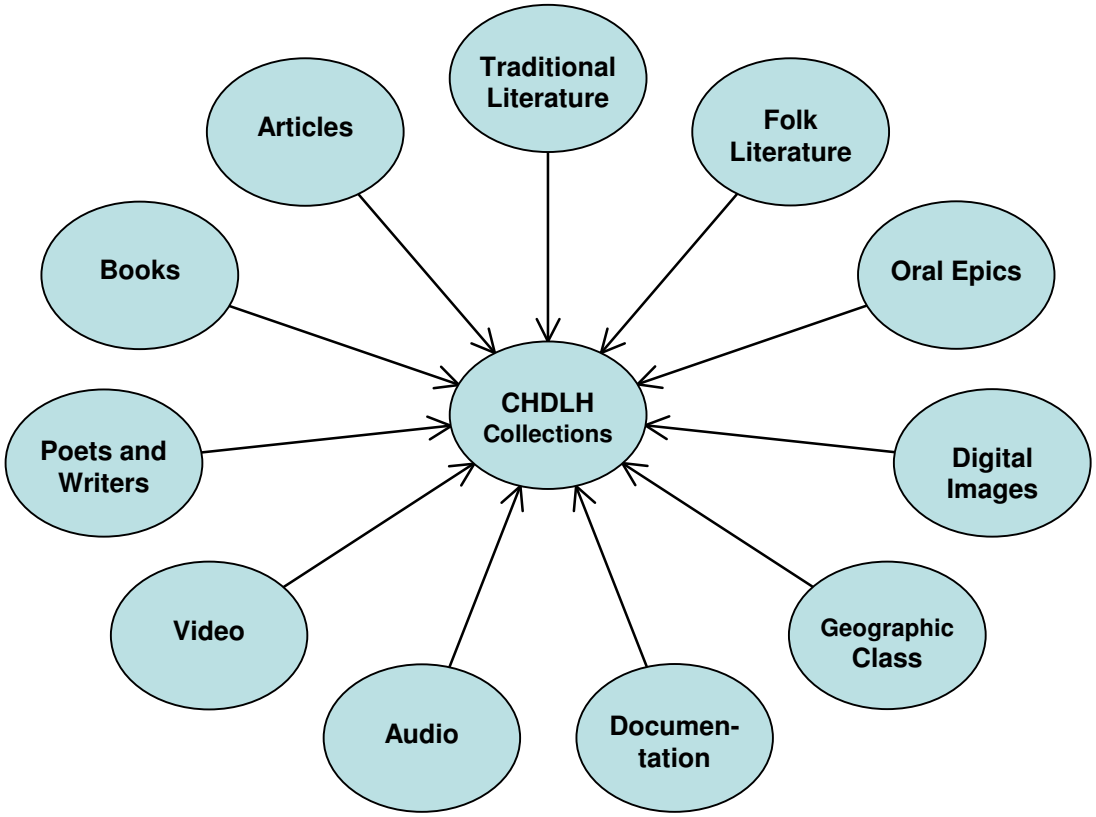


Fig. 4.21: CHDLH Collections with Multimedia Contents

4.5 National Databank on Indian Art and Culture (NDIAC) [www.ignca.nic.in/ndb_0001.htm]

After successfully completing two MCIT-supported projects by Cultural Informatics Lab of IGNCA in 10th Five Year Plan period, CIL becomes a vibrant executing agency in India for any kind of multimedia documentation or digitization project in the areas of Indian culture and heritage. The first two projects, undertaken by CIL, have some limitations in terms of their scopes. Most of digital resources created in DL-RICH project are not freely accessible online. On the other hand, digital resources available in CHDLH portal are in Hindi language that are seldom useful to the non Hindi-speaking worldwide communities. Thus, MCIT in association with ASI and IGNCA launched another digital library initiative, called 'National Databank on Indian Art and Culture' (NDIAC) in 2007 to be completed during the 11th Five Year Plan period.

Figure 4.22 presents a schematic structure of NDIAC project, where primary supporting partner of this project MCIT occupies top layer of the Figure followed by other key members of this project. Archaeological Survey of India (ASI) is the major stakeholder in this project as well as content provider, whereas CIL of IGNCA is the executing partner of the project.

This project envisages becoming one of the major authoritative sources of information on Indian art and culture. This project aims at enhancement of accessibility of Indian cultural resources using digital multimedia technology accessible from a single window. Figure 4.23 illustrates vision and objectives of NDIAC project as described in its proposal document. Although some of the objectives of this project is similar to other two projects of IGNCA described in the previous sections, NDIAC project avoids duplication or overlapping of information by way of selecting contents not covered in DL-RICH and CHDLH projects. NDIAC project has uniqueness in the way of supplementing the Ministry of Culture's newly launched National Mission on Monuments and Antiquities.

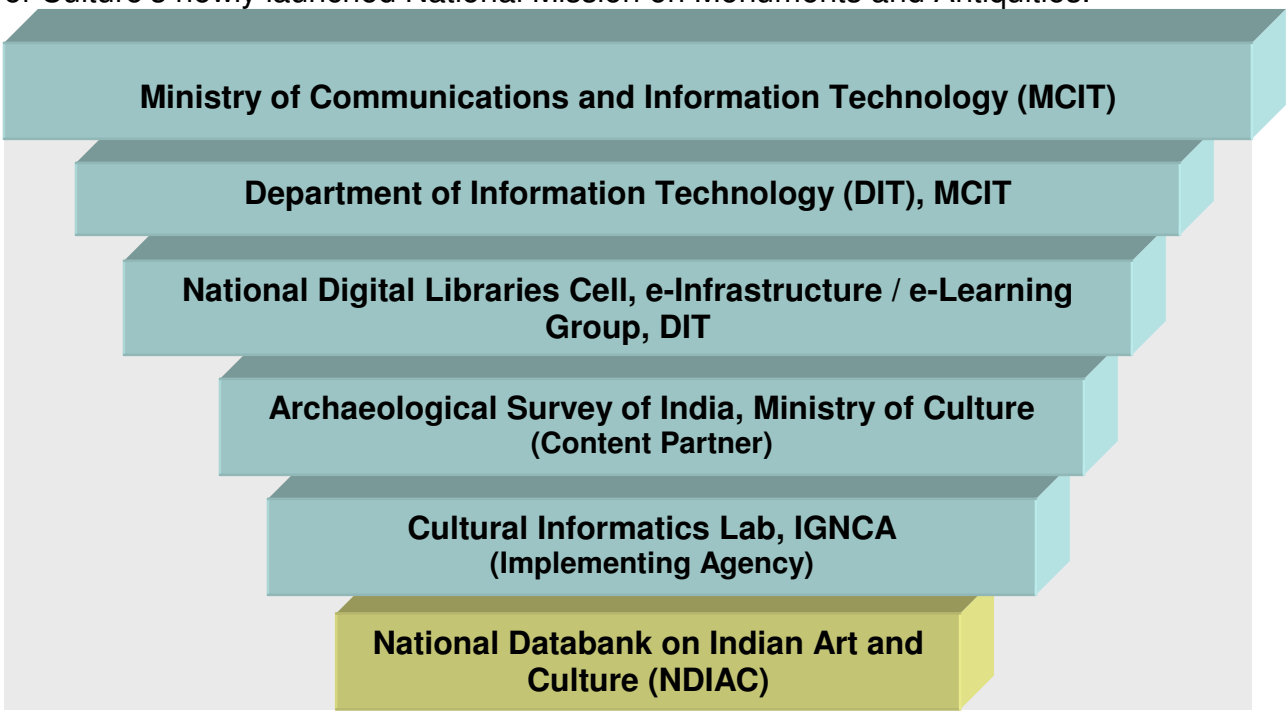
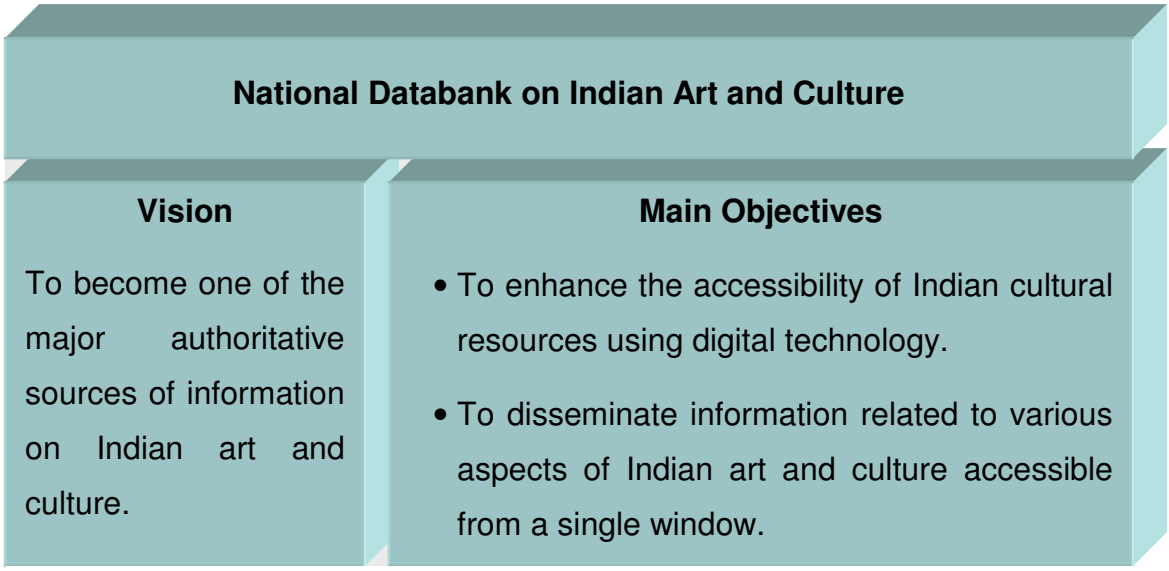


Fig. 4.22: Schematic Structure of NDIAC Project



Source: *About National Databank on Indian Art and Culture [www.ignca.nic.in/ndb_about.htm]*

Fig. 4.23: Vision and Objectives of NDIAC Project

4.5.1 Digitized Collections in NDIAC

NDIAC project incorporates textual as well as multimedia contents to be created or digitized by its partner institutions. Figure 4.24 indicates that NDIAC project has plan to have significant number of digitized books, digital images, audio recordings, video recordings and virtual walkthroughs on different aspects of Indian heritage, culture, arts and aesthetics.

Archaeological Survey of India (ASI) is the major content provider in this project as indicated in Figure 4.25. State archaeology departments will also play vital roles as content providers to this initiative. IGNCA is the implementing partner, digitization partner as well as content aggregator in this project.

NDIAC project supports retrospective digitization of ASI journal *Indian Archaeology - A Review*, starting from its first issue published in 1954 until year 2007. Many other important titles of ASI, presently out-of-print or not available in commercial channels, are also being digitized under the mandate of this project. Central Archaeological Library, a national level library under ASI, houses many rare books, art plates, original drawings, besides other antiquities. A significant number of rare books available with this library will be digitized in this project.

Visual archive in this project includes over 0.1 million digital photographs, photographic slides and negatives on various aspects of Indian art and culture collected from the various institutions and individuals. The photographic slides and negatives collected from different sources are digitized using appropriate digitization method and later stored in JPEG format. Respective groups of owners of the visual images share their digital copyrights with the project team and allow to make these visual images freely available through NDIAC Web portal for wide use by the researchers, archaeologists, art historians, students, and academics.

Virtual walkthrough is a kind of simulated interactive video tour to allow a user to virtually visit a heritage site. In this project five virtual walkthroughs will be created for worldwide access. Some of the selected sites for creation of virtual walkthroughs are inscribed on World Heritage List, maintained by UNESCO World Heritage Centre. The heritage sites selected for creation of virtual walkthroughs are:

- Brihadeshwara Temple, Thanjavur, Tamil Nadu
- Humayun's Tomb, New Delhi (world heritage site)
- Khajuraho, Madhya Pradesh (world heritage site)
- Rani ki Bagh, Patan, Gujarat (tentative world heritage site); and
- Martand Sun Temple, Anantnag, Jammu and Kashmir.

Table 4.7 indicates present status of the project as on 31st March 2008 vis-à-vis project targets. This Table also shows that creation of digital corpus is on its way and will meet the target hopefully by the end of project period.

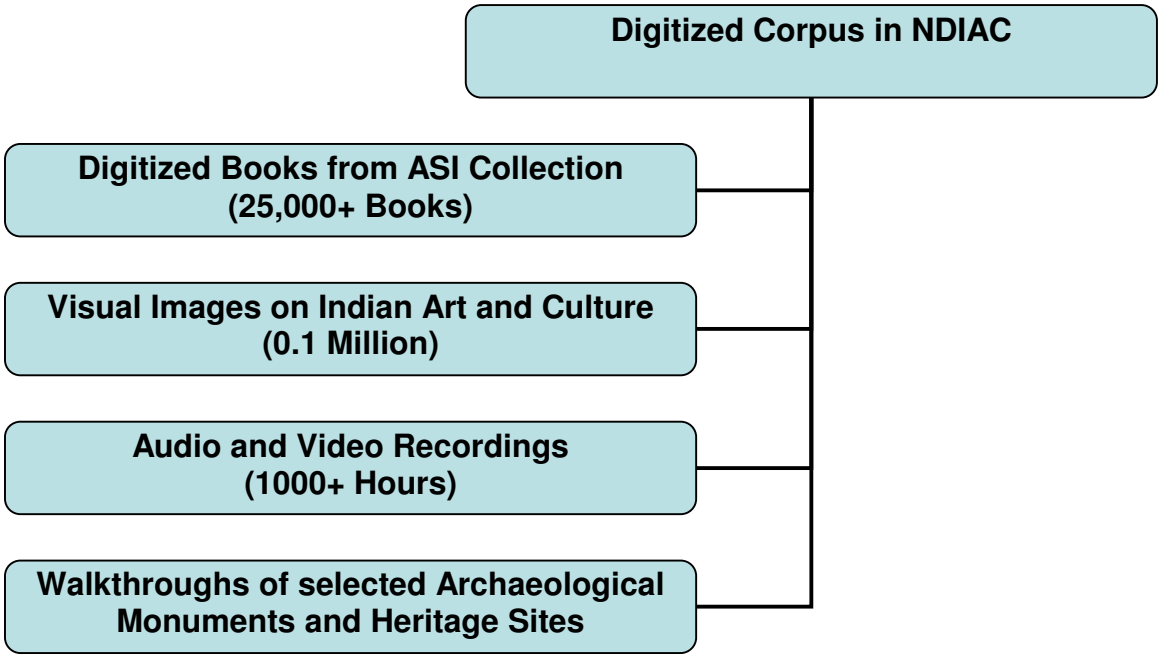


Fig. 4.24: Digitized Collections in NDIAC Project

Table 4.7: Digitization Progress of NDIAC Project

Activity	Project Targets	Achievement till 31st March 2008
Textual Pages	25000 books (5 Million pages)	1.2 Million pages (2157 Books)
Digital Images	0.1 Million	33000
Video (in hours)	300 hours	93 hours
Audio (in hours)	700 hours	35 hours
Walkthroughs	05	Nil
Source: *Targets achieved http://ignca.nic.in/ndb_targets.htm*		

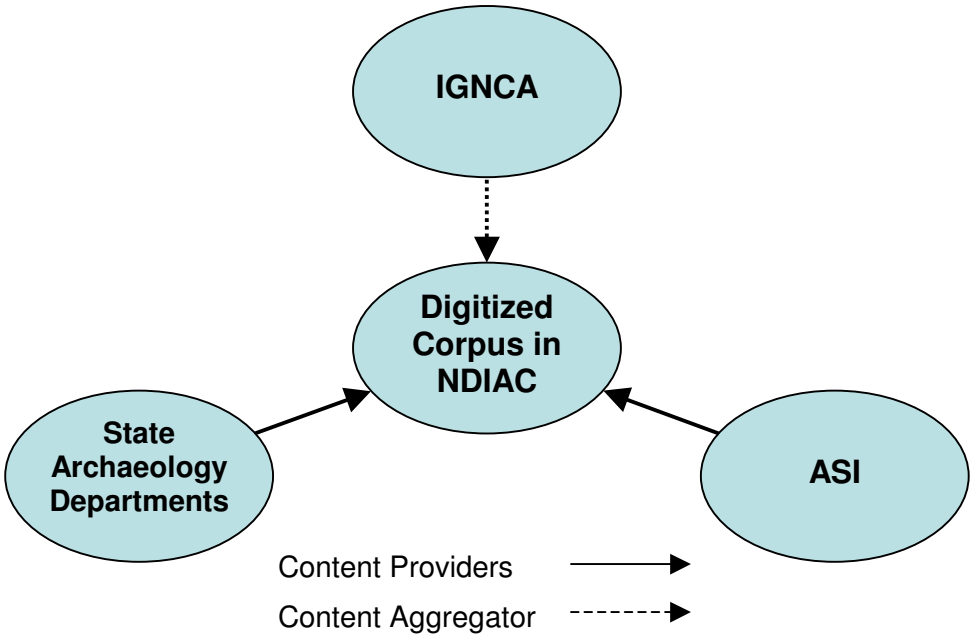


Fig. 4.25: Content Partners in NDIAC Project

4.5.2 Outreaching NDIAC Contents for Worldwide Audience

The multimedia cross-cultural contents emanated from this project will be disseminated through various channels. Corpus digitized from ASI resources will be uploaded on ASI's website for public access. There is also a plan to integrate NDIAC collections with existing open access digital library portals of IGNCA.

This project also envisages developing Travelling Multimedia Exhibition (TME) from the outputs of digitized collections. This TME will facilitate demonstration of the use of computer technologies for the preservation of Indian art and culture. Ministry of Culture or Ministry of Tourism can make use of this travelling exhibition for disseminating information on Indian heritage, arts and culture to the different corners of India and abroad. This way our government can attract a diverse audience to the cultural tourism in India.

Figure 4.26 identifies different modes of outreach to diversified audience for diffusing documented knowledge on the Indian heritage, art and culture.

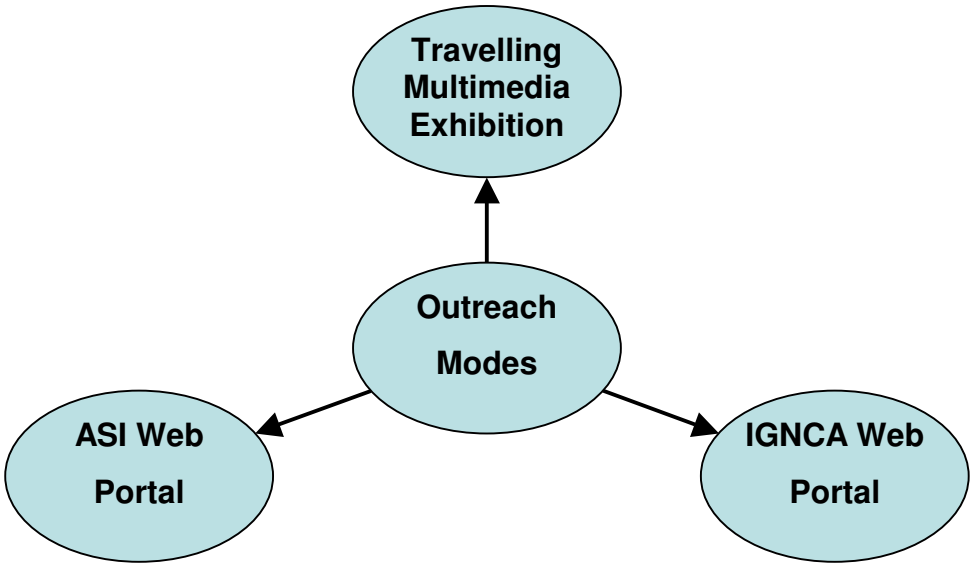


Fig. 4.26: Outreaching NDIAC Collections for Different Communities

This project is instrumental (i) to help researchers and archaeologists in searching and viewing the information of their interest, (ii) in bridging digital divide, and (iii) in promotion of cultural tourism in India, through freely accessible knowledge base on Internet created under this project.

4.6 National Mission for Manuscripts (NMM) [www.namami.org]

Old manuscripts are most significant sources of recorded information to portray socio-cultural and literary history of any civilization. Before invention of printing machine, manuscript was the only medium for recording and communicating human knowledge and wisdom. The manuscript tradition has carried forward from generations to generations and centuries to centuries. Indian civilization is one of the oldest civilizations in the world and thus India perhaps possesses one of the oldest and largest collections of manuscripts in the world. These manuscripts are inscribed on different kinds of materials such as birch bark, palm leaf, cloth and paper. Indian manuscripts are available in different scripts and in different languages. A variety of institutions such as libraries, museums, archives,

monasteries, madrasas, temples, mutts as well as private homes are custodians of Indian manuscripts. It is estimated that about five million Indian manuscripts are available in India and another 0.3 to 0.5 million in other countries [1] (Department of Culture, 2002). Indic researchers and scholars could know only a small portion of manuscripts (about 20-25%) from the secondary sources such as printed catalogues, library catalogues and manuscript surveys out of total collections of Indian manuscripts. This is because untraceable treasures of manuscripts are hidden in private homes, religious trusts, houses of worship and defunct public bodies. On the other hand, manuscripts inscribed on biodegradable materials and may perish due to natural decay after certain period. Manuscripts are often found to have been neglected for decades and in very poor physical condition – insect ridden, fungus infected or brittle, fading and fragile. Most of the manuscript collections in India are neither scientifically preserved nor properly documented through printed catalogues or library catalogues or registers.

In this situation, Government of India launched National Mission for Manuscripts (NMM) in February 2003 to locate, enumerate, preserve and describe all of the Indian manuscripts in India and abroad. The objectives for undertaking these tasks are to document and preserve millions of neglected manuscripts, enhance their access, spread awareness about this cultural inheritance and encourage their use for educational and research purposes and lifelong learning.

National Mission for Manuscripts (NMM) is a time-bound programme of Government of India, initially established for execution during 10th Five Year Plan period. Department of Culture is implementing this mission-mode programme through establishment of institutional frameworks and national network of manuscript repositories across the country. Indira Gandhi National Centre for the Arts is the national nodal agency for this mission. IGNCA provides all kinds of institutional, functional and technical support to NMM including national office space.

[1] Project Document: National Mission for Manuscripts. New Delhi: Department of Culture, 2002.

NMM established a network of manuscript repositories to achieve its different objectives. Network members are categorically differentiated based on their assigned responsibilities. Network members are designated as: (i) Manuscript Resource Centre (MRC), (ii) Manuscript Conservation Centre (MCC), (iii) Manuscript Partner Centre (MPC), (iv) Manuscript Conservation Partner Centre (MCPC), and (v) Manuscript Digitization Centre (MDC). A manuscript repository can perform one or more responsibilities of designated centres. MRC is the backbone of national documentation tasks of the Mission, whereas other kinds of centres serve specific purposes of the Mission. Figure 4.27 provides a schematic structure of the Mission, where Ministry of Culture (erstwhile Department of Culture during launching of the Mission) is placed in top layer as a parent body of the Mission, followed by national nodal agency IGNCA. Manuscripts digitization centres are obtained from other four type partner centres of NMM.

For smooth functioning of the Mission, Ministry of Culture formed several committees as mentioned in Figure 4.28. National Empowered Committee is the advisory body of the Mission that provides visionary directions to the Mission. Executive Committee is responsible for execution and implementation of planned activities across the country. Finance Committee is responsible for professional handling of financial resources for proper execution of Mission's activities. Project Monitoring Committee plays monitoring and evaluation role, which is very important for a time-bound programme in reducing any deviation from predefined goals.

National Mission for Manuscripts has some comprehensible broad objectives. These objectives also clearly convey a plan of actions of NMM to address the challenges faced by Indian society to deal with undocumented and unpreserved manuscript collections across India. In its objectives NMM states that it will undertake national level survey and post-survey to locate rare manuscript collections covering major districts in every state of India. The information collected from national surveys will be documented, catalogued and stored in National Electronic Register of Manuscripts. This is an online catalogue database

of Indian manuscripts, presently having more than 1.5 million records. After documenting all revealed data on Indian manuscript collections, NMM targets physical preservation as well as digital preservation of most significant as well as most endangered pieces of manuscripts. Figure 4.29 illustrates broad and digitization objectives of NMM.

Digitization of manuscripts is one of the major intervention areas of the Mission. NMM envisages that digitization will improve access to significant Indian manuscript collections. In its digitization objectives, NMM states that digitization helps in long-term preservation of manuscripts. Digitization also helps in reducing physical handling of rare and fragile original manuscripts. Digital surrogates of manuscripts are much easier to share and use than a unique copy of original manuscripts archived in special condition. National Manuscripts Library, having a digital library component, will store digitized collections of Indian manuscripts and eventually will be accessible online. NMM also envisaged to create standards and procedures for digitization of manuscripts, which NMM already documented in its booklet titled *Guidelines for Digitization of Manuscripts*, published in 2005.

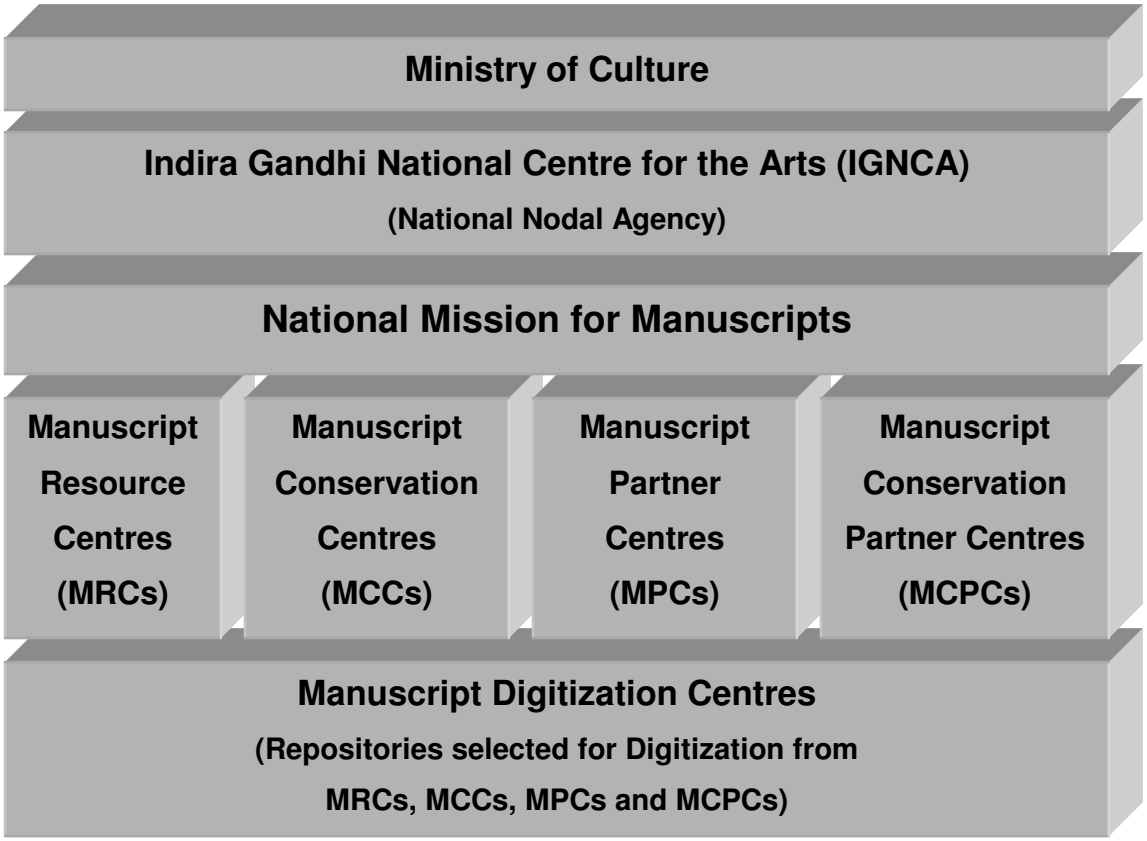


Fig. 4.27: Schematic Structure of the National Mission for Manuscripts

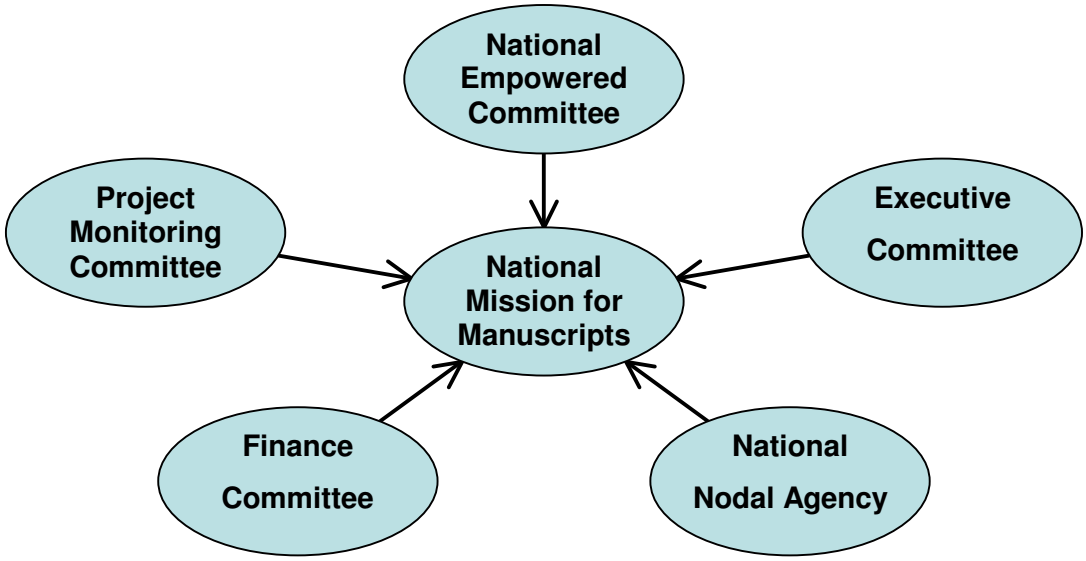


Fig. 4.28: Committees Governing the National Mission for Manuscripts

National Mission for Manuscripts

Broad Objectives

- To locate manuscripts through national level survey and post-survey.
- ***To document each and every manuscript and manuscript repository in a National Electronic Database making this the largest database on Indian manuscripts in the world.***
- To conserve manuscripts incorporating both modern and indigenous methods of conservation and training a new generation of manuscript conservators.
- To train the next generation of scholars in various aspects of manuscript studies such as languages, scripts, critical editing, cataloguing of texts and conservation of manuscripts.
- ***To promote access to manuscripts by digitizing the rarest and most endangered manuscripts.***
- To promote access to manuscripts through publication of critical editions of unpublished manuscripts and catalogues.
- To facilitate public's engagement with manuscripts through lectures, seminars, publications and other outreach programmes.

Digitization Objectives

- To preserve the original manuscripts for posterity.
- To promote access and usage for scholars and researchers, without tampering with original copies.
- To reduce handling of very fragile or frequently used original manuscripts.
- To preserve in a longer lasting medium and to create a back up copy with as close a resemblance to the original as possible.
- To make accessible this vast archival resource to users through the National Manuscripts Library to be eventually set up by the Mission.
- To create a digital library as a resource base of the digitized copies of some of the significant manuscript collections of the country.
- To create standards and procedures for digitization of manuscripts.

Fig. 4.29: Broad Objectives and Digitization Objectives of NMM

4.6.1 National Documentation Projects of NMM

NMM undertakes national surveys in each state and union territory for locating manuscripts in both known and unknown, private and public, catalogued and non-catalogued collections through a standard questionnaire form as shown in Annex 3. After completion of national surveys, NMM undertakes extensive post-survey exercises to document each manuscript in Manus Data Sheets (Annex 4).

National survey addresses documentation of whole manuscript collection of an institution or individual, whereas post-survey addresses documentation of detailed information of each manuscript available with surveyed institution or individual.

In the exercises of national surveys and post surveys, state and district coordinators provide local support in enumeration of data in structured formats. Usually, state and district coordinators are drawn from existing Manuscript Resource Centres of NMM.

During post-survey, local coordinator receives filled in Manus Data Sheets from the surveyors. Manus Data Sheet describes each manuscript with detailed information such as title, author, commentary, script, name of repository, name of scribe, date of manuscript, number of folios and pages, and other relevant information. Detailed information collected using Manus Data Sheets are then entered into decentralized national online catalogue of manuscripts using e-Granthavali software. This catalogue database is also known as National Electronic Register of Manuscripts. After cataloguing all surveyed manuscripts, NMM will select most significant manuscripts for digitization. Then the process of digitization will be started in selected centres of NMM.

Figure 4.30 illustrates the lifecycle of documentation activities in NMM for Indian manuscripts. Activities carried out in the first phase of lifecycle are national survey and post-surveys across the country. In the second phase of lifecycle, NMM incorporates surveyed data in the National Electronic Register of Manuscripts – an online catalogue of manuscripts. In the third phase of lifecycle, NMM undertakes digitization activities and incorporates digitized collections in

the National Manuscripts Library – hybrid library having both online digital library as well as offline physical and digital collections housed in IGNCA premises.

Figure 4.31 more specifically indicates that NMM maintains two types of online databases, namely 'Kritisampada: National Database of Manuscripts' and National Manuscripts Library. Kritisampada is the Web-based search engine of National Electronic Register of Manuscripts, presently available from the NMM portal (Arora, 2007). Digitized collections of manuscripts have been incorporated in the National Manuscripts Library, which is a kind of hybrid library having both physical and digital collections. Digital collections of National Manuscripts Library are now available on intranet only at the IGNCA premises.

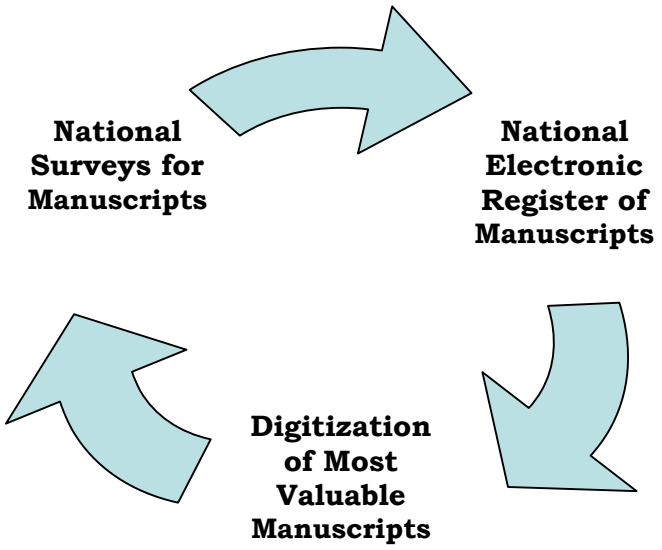


Fig. 4.30: Life Cycle in Documentation of Indian Manuscripts

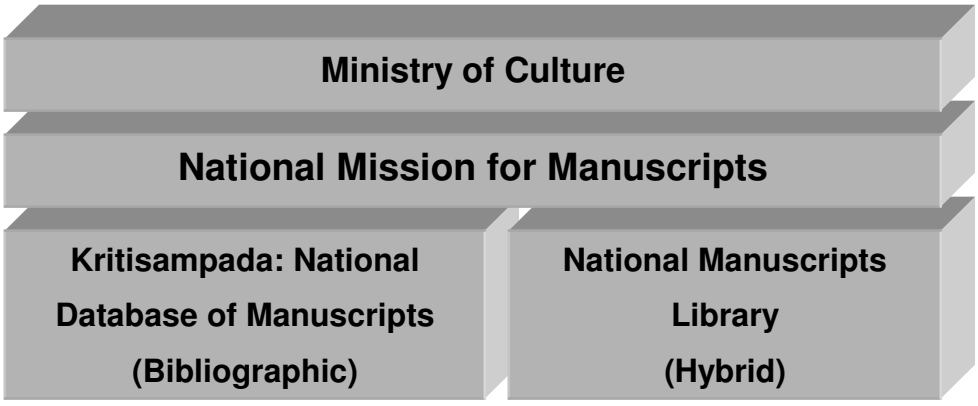


Fig. 4.31: Online Databases of NMM

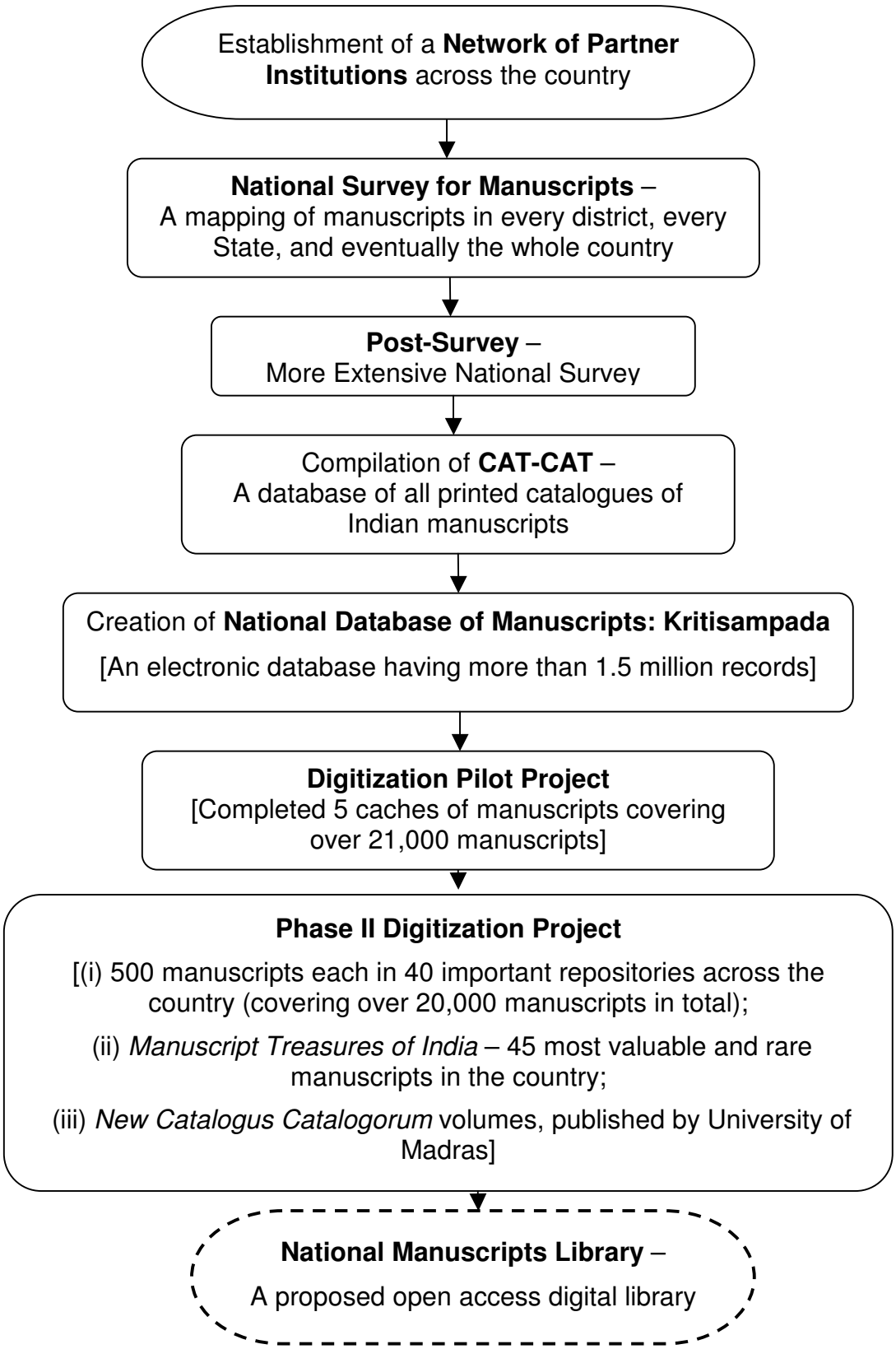


Fig. 4.32: Workflow of NMM Activities towards the Creation of National Manuscripts Library

4.6.2 Workflow of Documentation NMM Activities

Although, NMM initiated as a time-bound mission mode programme, it started execution of many planned intervention activities within a short-time span. NMM established a network of partner institutions across the country with a specific role for an institution such as Manuscript Resource Centre (MRC), Manuscript Conservation Centre (MCC), Manuscript Partner Centre (MPC), or Manuscript Conservation Partner Centre (MCPC).

NMM undertakes national survey for manuscripts - mapping of manuscripts in every district, every state, and eventually the whole country involving mainly MRCs and MPCs as state and district coordinators. After completion of national level surveys to document manuscript heritage in Indian states, NMM undertakes more extensive national post-survey to obtain bibliographic information of every manuscript, revealed in national level surveys. For the national survey, a structured questionnaire is used to describe manuscript holdings of each surveyed institution or individual. For the post-survey, a structured Manus Data Sheet is used to describe individual manuscript.

Printed catalogues describe institutional holdings of manuscript collections in India and abroad. NMM collates all available catalogues of Indian manuscripts, published or unpublished, for the compilation of CAT-CAT – a union catalogue database for Indian manuscripts.

NMM maintains national electronic catalogue of manuscripts incorporating surveyed manuscript collections as well as CAT-CAT records. A derivative of this catalogue database is the 'Kritisampada: National Database of Manuscripts' presently available on NMM Web portal describing more than 1.5 million manuscripts available across India.

NMM undertakes digitization projects in two phases. The first phase is known as digitization pilot project that digitized five caches of manuscripts covering over 21,000 manuscripts. In the second phase of digitization project, digitization is carried out in forty important repositories across the country covering over 20,000 manuscripts in total. In the later stage of this phase, the project will cover

digitization of (i) Manuscript Treasures of India – 45 most valuable and rare manuscripts in the country; and (ii) New Catalogus Catalogorum volumes, published by University of Madras. After completion of ongoing digitization projects, it is proposed to provide open access to digitized collections through National Manuscripts Library. Presently National Manuscripts Library is available at IGNCA premises only for national consultation having both digitized and physical collections. The physical manuscript collections are built up from the donations of manuscripts voluntarily given by individuals and public trusts.

4.6.3 Manus Granthavali – the Electronic Cataloguing Tool for Indian Manuscripts

Manus Granthavali, the National Electronic Register application software developed by NIC, is based on the globally accepted Unicode standards. It supports data entry in thirteen Indian languages. The reasons for choosing Unicode standards are its global acceptance, easier localization of application, as well as improved multilingual text processing on the data related to Indian manuscripts. This application software allows the content to be transliterated from one language to another with great ease.

This client server application has been built on the .NET platform with Microsoft SQL Server™ 2000 (Microsoft India, 2005).

This software is based on Dublin Core metadata standards that are used in describing online documents. All the data fields in Manus Data Sheets are decoded into the qualified Dublin Core metadata elements.

Figure 4.33 shows software brochure indicating its specifications, salient features, and IT requirements for implementation.

This software is installed in all manuscripts resource centres (MRCs) and manuscript partner centres across India. Some of them are state coordinators for the national surveys and post-surveys. Thus, surveyed and filled up Manus Data Sheets are entered into the National Electronic Register at the state coordinating centre of the respective state using this software. Later, the entered catalogue

records are sent to NMM office in New Delhi for integrating into the national database. NIC provides technical support to the state coordinators of national surveys, MRCs and MPCs for smooth implementation of Manus Granthavali at their locations. It is estimated that approximately 5 million manuscripts have to be catalogued using this software. Presently, Web-based National Database of Manuscripts: Kritisampada provides information on about 1.5 million manuscripts.

Fig. 4.33: Brochure of Manus Granthavali – the Software Used for Creation of National Electronic Register of Manuscripts

4.6.4 Kritisampada: National Database of Manuscripts

The 'Kritisampada: National Database of Manuscripts' is an online catalogue of about 1.5 million manuscripts, recorded across the country through the national level surveys and post-surveys. This database contains bibliographic records of Indian manuscripts, obtained from different memory institutions, religious institutions, private bodies and individuals through an instrument called Manus Data Sheet. This database is derived from the National Electronic Register of Manuscripts, maintained by NMM.

This online database can be searched by seven metadata elements such as title of manuscript, author of manuscript, language, material type, state of India, subject heading, and script. Free text search is possible for the metadata elements – author and title. For any other metadata element, the search engine populates full list of recorded information from that category and user has to select from that list. For example, language metadata element has a list of 365 languages. Similarly, 20 types of materials are recorded in this database.

Figure 4.34 illustrates different search options available with this online database. Out of all these metadata elements indicated in this Figure, individual repository can be browsed on the basis of state, then district. The user has to first select a state/ UT of India, followed by district and name of repository. When a repository is selected, then the user can view full list of manuscripts available with that repository. When the user pinpoints a particular manuscript, full bibliographic detail of that manuscript is displayed on screen.

As indicated in Figure 4.34, Kritisampada database can be searched by type of material and name of state/ union territory of India. Table 4.8 provides a list of available material types recorded in Kritisampada database. As indicated in this Table, manuscripts are inscribed on as wide as twenty types of materials. Out of them, paper, palm leaf, cloth, birch-bark and scroll manuscripts are commonly available in Indian manuscript repositories. This Table also provides a list of Indian states and union territories covered in this database. This database

recorded manuscripts information from three union territories out of seven UTs/NCT and nineteen states of India out of twenty-eight states.

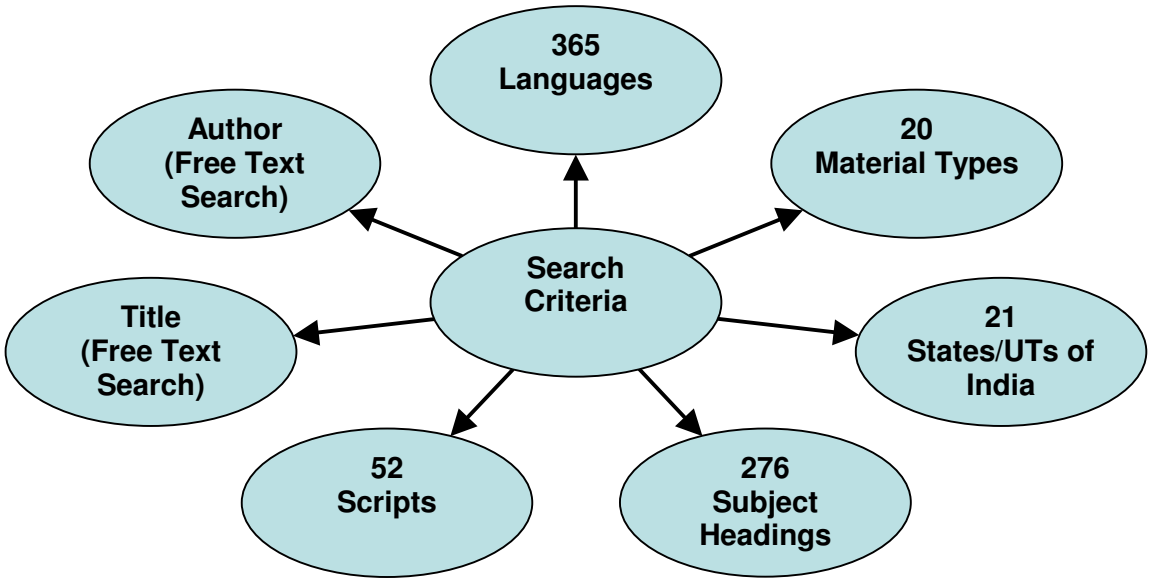


Fig. 4.34: Searching and Navigation Options available at Kritisampada

Table 4.8: Available Types of Materials and Names of States

Types of Materials		**States/ UT Coverage**	
Bamboo Leaves	Papers	**States**	Maharashtra
Bhoj Patra	Parchments	Andhra Pradesh	Orissa
Birch-bark	Sanchipats	Assam	Punjab
Clay tablets	Scrolls	Bihar	Rajasthan
Cloth	Tamra Patra	Chhattisgarh	Tamil Nadu
Deer Skin	Tortoise shell	Gujarat	Uttar Pradesh
Ivory Covers	Tulapats	Haryana	Uttarakhand
Leather	Wooden Beads	Jammu & Kashmir	West Bengal
Palm Leaf	Wooden Covers	Jharkhand	**UTs/ NCT**
Pamera Leaves	Others	Karnataka	Chandigarh
		Kerala	Delhi
		Madhya Pradesh	Pondicherry

4.6.5 Pilot Digitization Project

As national surveys and post-surveys recorded an ocean of manuscript collections, it became very difficult for NMM to digitize all manuscripts available with its partner institutions. Thus, NMM planned a pilot digitization project to digitize initially only those manuscripts with scientific, artistic, spiritual or heritage values and those that are endangered. The pilot project is focused on creating high-resolution and low-resolution images, which can either be printed or viewed over the Internet.

In this phase, over 21,000 manuscripts were digitized in five states covering five significant caches of manuscripts. Figure 4.35 provides a list of caches of manuscripts digitized in this phase. Figure 4.36 indicates a list of digitization agencies involved in providing digitization services for successful completion of this pilot digitization project.

In this phase, Cultural Informatics Lab of IGNCA digitized manuscript collections of Iqbal Library in Srinagar covering endangered manuscripts of the Valley. Siddha Manuscripts, related to Siddha medicine practised mainly in south India, are archived in different repositories in Tamil Nadu, including one in the Institute of Asian Studies in Chennai. The Tamil Medical Manuscript Collection of the Asian Studies is inscribed on the Memory of the World Register in 1997 that implies that this collection is globally significant. In the digitization pilot project of NMM these Siddha manuscripts were digitized by Mahabharata Samshodhana Pratishthanam in Bangalore, a group set up by software engineers working with Wipro India Limited. This group also developed a software called ‘Vyasa’ for the digitization purposes.

In this phase National Informatics Centre in Bhubaneswar digitized illustrated manuscripts of Orissa, whereas the Centre for Digital Imaging Technology in Trivandrum digitized Kutiyattam Manuscripts from Kerala.

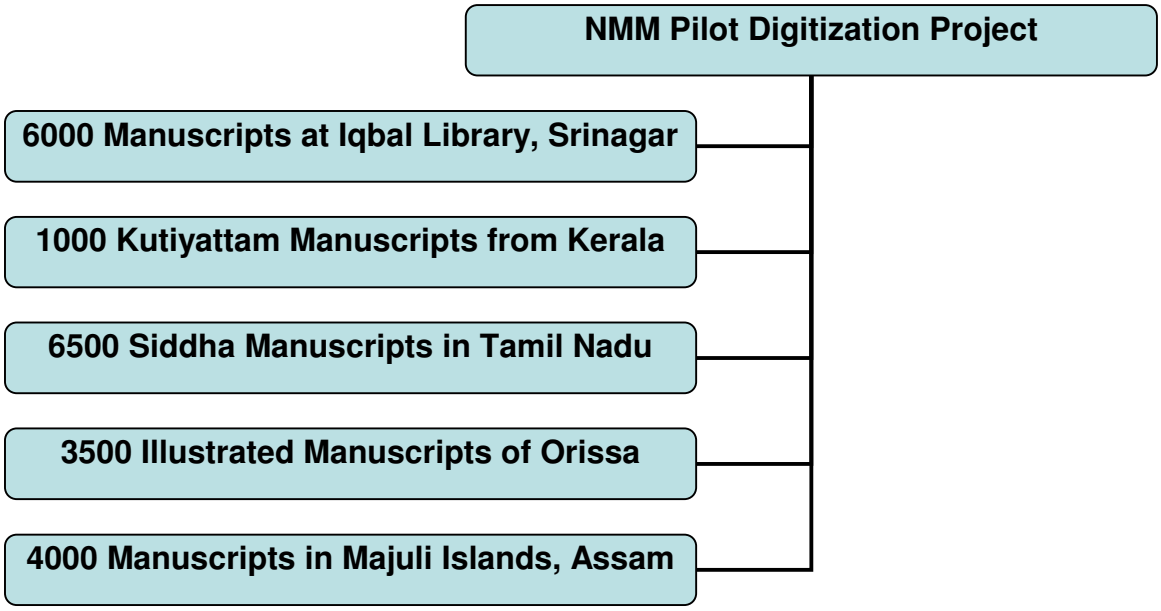


Fig. 4.35: Pilot Digitization Project Completed by NMM

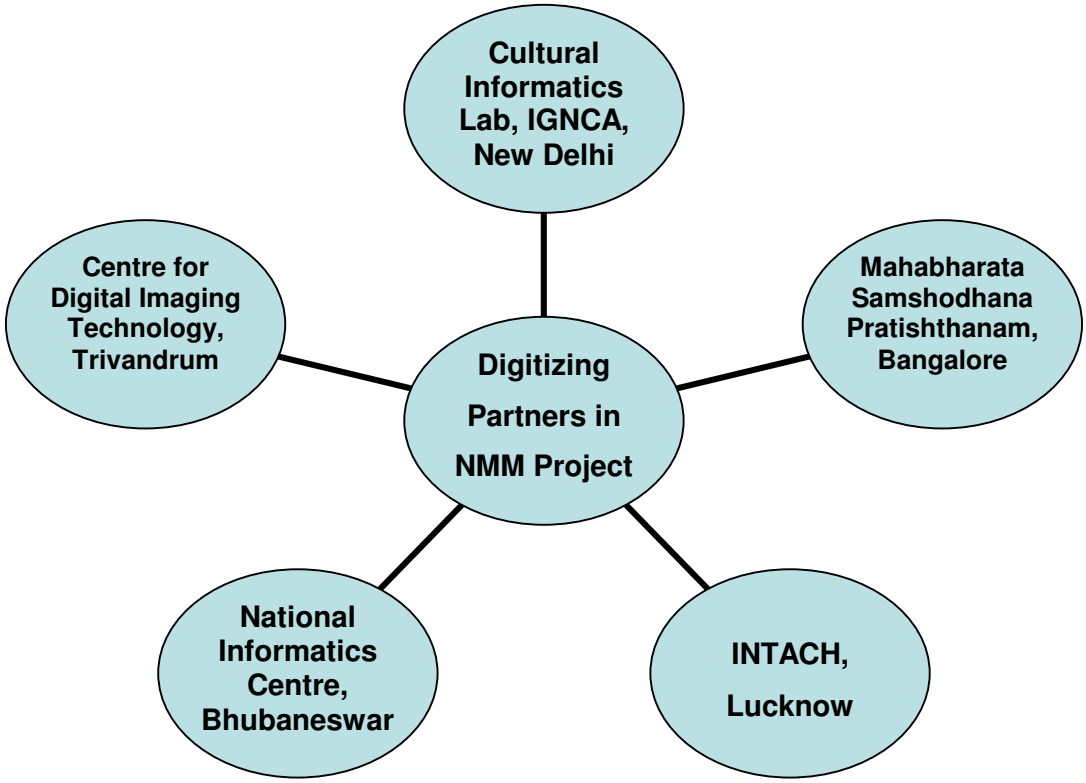


Fig. 4.36: Digitization Partners of NMM in Phase I Project

4.6.6 Guidelines for Digitization of Manuscripts

Just before the pilot digitization project, NMM published a guideline booklet titled 'Guidelines for Digitization of Manuscripts' based on international standards, guidelines and best practices. These guidelines were applied and tested in the first phase of digitization project. As second phase is implemented in more decentralized manner, these guidelines will help manuscript digitization centres to comply with output quality of digitized manuscripts including standards related to metadata elements. For instance, each folio of manuscript is scanned with 300dpi spatial resolution and stored in TIFF format. A cleaned image can also be produced from the original image file and is stored in another file in TIFF format with 300dpi resolution. An access image file is also generated from the cleaned image file that is stored in JPEG format suitable for Web access. In short image formats for scanned images are:

(i) Master Image – TIFF Format
(ii) Clean Image – TIFF Format
(iii) Access Image – JPEG Format
(iv) Thumbnail Image – JPEG Format

This guideline also suggests a list of metadata elements set, similar to metadata elements used in the Manus Data Sheet or Manus Granthavali software application.

4.6.7 Phase II Digitization Project

After successfully completion of Digitization Pilot Project, NMM planned for second phase digitization project covering fifty important repositories across the country. Each participating repository in second phase digitization project will identify 500 most significant manuscripts from its collection for digitization. In later stage, number of repositories to be covered in the second phase project is reduced to forty due to time constraints. Thus, this phase covers about 20,000

manuscripts in total for digitization. Figure 4.37 shows zone-wise distribution of manuscript digitization centres across India in the second phase digitization project. Table 4.9 provides a list of manuscript repositories selected for carrying out second phase manuscript digitization project. This Table identifies the existing role of each repository that eventually becomes a manuscript digitization centre of NMM. The Table also indicates that most of the repositories selected for second phase digitization project presently act as manuscript resource centres of NMM and a few of them play other roles. Figure 4.38 indicates a wide coverage of Indian states and union territories in the first and second phases of digitization projects of NMM.

In addition to forty selected repositories as indicated in Figure 4.37, this phase also covers three more important collections, namely,

(i) Jain Manuscripts: INTACH, Lucknow has already started digitization of this important collection covering about one million folios.

(ii) Vijnananidhi: Manuscript Treasures of India – NMM identified 45 most valuable and rare manuscripts in the country. For this purpose NMM published an illustrative book with the same title with descriptive catalogue of each identified manuscript (Paliwal, 2007).

(iii) New Catalogus Catalogorum volumes: NMM is digitizing the existing volumes of the New Catalogus Catalogorum for preserving the valuable data in these reference sources. NMM is collaborating with the publisher, University of Madras, for this purpose (NMM, 2007).

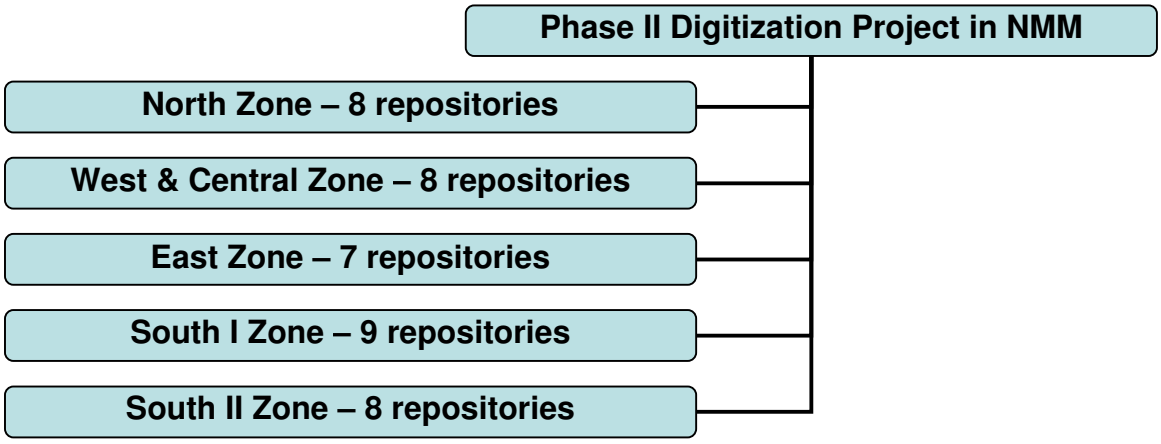


Fig. 4.37: Phase Two Digitization Project Undertaken by NMM

Table 4.9: Manuscript Repositories Selected for Phase II Project

Name of Repository	City	State	Existing Role(s)
North Zone			
Vrindavan Research Institute	Vrindavan	Uttar Pradesh	MRC[1], MCC[2]
Aligarh Muslim University	Aligarh	Uttar Pradesh	MCPC[3]
Rampur Raza Library	Rampur	Uttar Pradesh	MRC, MCC
Himachal Academy of Art, Culture & Language	Shimla	Himachal Pradesh	MRC
Vishveshvaranand Vedic Research Institute	Hoshiarpur	Punjab	MRC, MCC
Uttaranchal Sanskrit Academy	Haridwar	Uttarakhand	MRC
Central Institute of Buddhist Studies	Leh	Jammu & Kashmir	MRC
Shri Ranbir Sanskrit Research Institute	Jammu	Jammu & Kashmir	-
West + Central Zone			
Rajasthan Oriental Research Institute	Jodhpur	Rajasthan	MRC
Institute of Rajasthan Studies	Udaipur	Rajasthan	MPC[4]
Lalbhai Dalpatbhai Institute of Indology	Ahmedabad	Gujarat	MRC
Shivaji University	Kohlapur	Maharashtra	MPC

Cont…

[1] Manuscript Resource Centre
[2] Manuscript Conservation Centre
[3] Manuscript Conservation Partner Centre
[4] Manuscript Partner Centre

Cont…

Name of Repository	City	State	Existing Role(s)
Shree Forbes Gujarati Sabha	Mumbai	Maharashtra	MPC
Scindia Oriental Research Institute	Ujjain	Madhya Pradesh	MRC, MCC
Dr. Harisingh Gour University	Sagar	Madhya Pradesh	MRC
Kunda-Kunda Jnanpitha	Indore	Madhya Pradesh	MRC
East Zone			
Kameshwar Singh Darbhanga Sanskrit University	Darbhanga	Bihar	MRC
Nav Nalanda Mahavihara	Nalanda	Bihar	MRC
Sri Dev Kumar Jain Oriental Research Institute	Arrah	Bihar	MRC, MCC
Manipur State Archives	Imphal	Manipur	MRC, MCC
Krishna Kanta Handiqui Library	Guwahati	Assam	MRC
Calcutta University	Kolkata	West Bengal	MRC, MCC
Sanskrit Sahitya Parishad	Kolkata	West Bengal	MCPC
South I Zone			
French Institute of Pondicherry	Pondicherry	Pondicherry	MRC
Institute of Asian Studies	Chennai	Tamil Nadu	MPC

Cont…

Cont…

Name of Repository	City	State	Existing Role(s)
Government Oriental Manuscripts Library (GOML)	Chennai	Tamil Nadu	MCPC
Sarasvati Mahal Library	Thanjvur	Tamil Nadu	MRC, MCC
Tamil University	Thanjvur	Tamil Nadu	MPC
C.P. Ramaswamy Aiyar Foundation	Chennai	Tamil Nadu	MPC
Muhammadan Public Library	Chennai	Tamil Nadu	-
Calicut University	Calicut	Kerala	-
Oriental Research Institute	Trivandrum	Kerala	MRC
South II Zone			
Oriental Research Institute	Tirupati	Andhra Pradesh	MRC, MCC
Salarjung Museum	Hyderabad	Andhra Pradesh	MCC
Idara-e-Adabiyat-e-Urdu	Hyderabad	Andhra Pradesh	MPC
Oriental Manuscripts Library and Research Institute (OMLRI)	Hyderabad	Andhra Pradesh	MRC
National Institute of Prakrit Studies	Shravanbelgola	Karnataka	MRC
Keladi Museum	Keladi	Karnataka	MRC
Kannada University	Hampi	Karnataka	MRC
Oriental Research Institute	Mysore	Karnataka	MRC

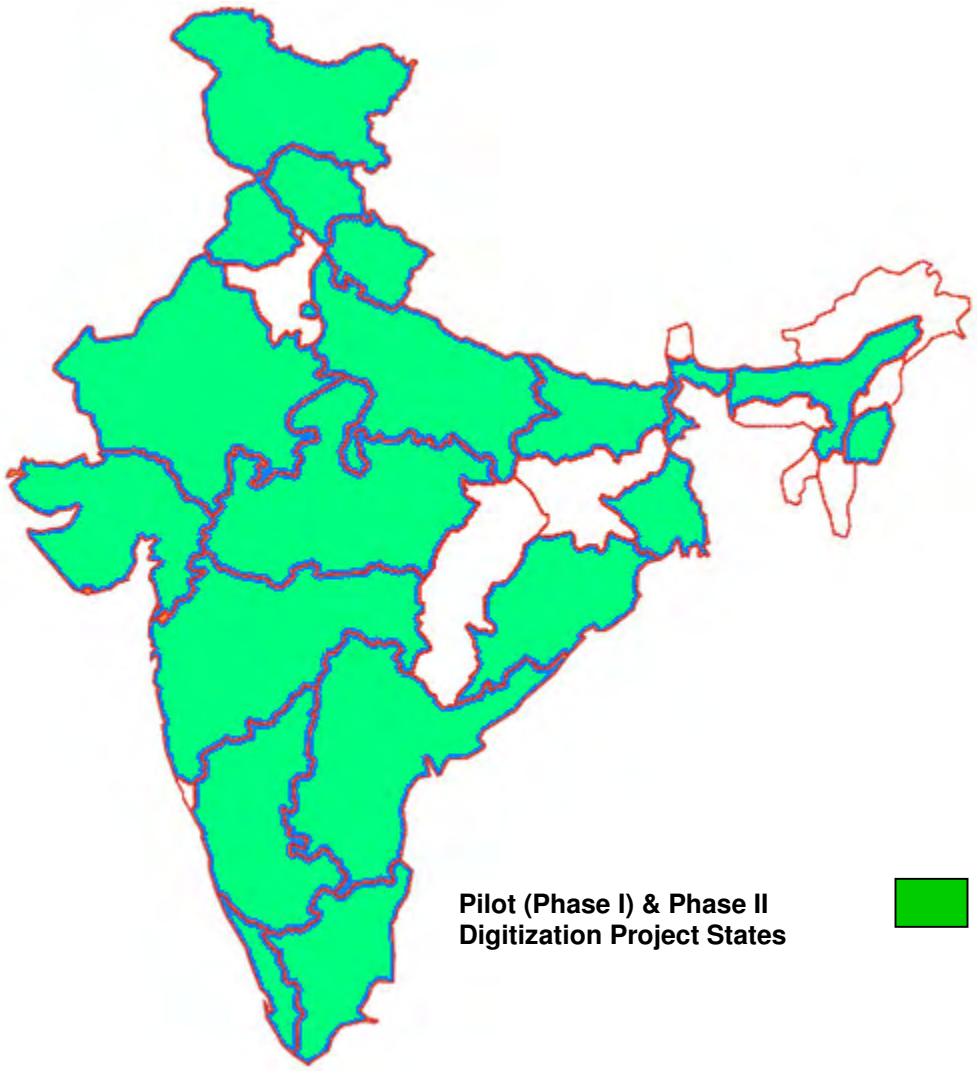


Fig. 4.38: NMM Digitization Project States in Two Phases

4.7 Some Observations

In this chapter only two national level digitization programmes for documentary heritage collections have been analysed. In India, few other national and institutional level initiatives for digital preservation of documentary heritage collections are worth mentioning. Some of the important exemplary initiatives are: (i) *Muktabodha Online Digital Library* of Muktabodha Indological Research Institute; (ii) *Traditional Knowledge Digital Library* of Council of Scientific and Industrial Research (CSIR) and Department of Ayurveda, Yoga & Naturopathy,

Unani, Siddha and Homoeopathy (AYUSH); (iii) *Archives of Indian Labour* of V. V. Giri National Labour Institute & Association of Indian Labour Historians; (iv) *Online Digital Collections* of Khuda Bakhsh Oriental Public Library; (v) *Down Memory Lane* of National Library of India, besides few others.

Indian digitization initiatives carefully adopted international standards and procedures. They have tested some of these standards and procedures to derive certain procedures best suitable for Indian environment and Indian documentary heritage collections. Both organizations described in this chapter – IGNCA and NMM – have adopted international standards and documented best practice guidelines for their partner institutions.

The Indian organizations undertake digitization activities for documentary heritage collections either on mission mode or project mode. For both the cases, successful implementation of a digitization project may lead to proposing another digitization project to cover another block of whole mass of documentary heritage resources. In the mean time, the institutions involved in digitization projects including partner institutions obtain a level of capability maturity that helps in capacity development of a country.

The digitization projects, described in this chapter, have contributed towards achieving multicultural and cross-cultural dialogs in a democratic society, in addition to making the endangered documentary resources digitally available. Digitization of documentary heritage collections in a culturally rich and diverse country is a major challenge to the digital library professionals and policymakers, due to nature of vastness versus available financial resources and institutional frameworks. Thus, scaling up is real concern in India that needs to involve all possible stakeholders as well as end users.

4.8 References

Arora, A. (2007). Online catalogue of one million manuscripts – Kritisampada. *Kriti Rakshana*, 2(4), 3-5.

Department of Culture (2002). *Project document: National Mission for Manuscripts*. New Delhi: Department of Culture.

Gaur, R.C. (2006). *Digital libraries in India with special reference to experiences at IGNCA, New Delhi, India*. Retrieved from http://publikationen.ub.uni-frankfurt.de/volltexte/2006/2445/.

Indira Gandhi National Centre for the Arts (2003). CoIL-Net Centre: Cultural Heritage Digital Library in Hindi. *Viswabharat @TDIL*, 9, 47-55.

Indira Gandhi National Centre for the Arts (2007). *Cultural informatics: information brochure of Cultural Informatics Lab*. New Delhi: IGNCA.

Jha, P. (2005). Digital Library @ IGNCA. *Viswabharat @TDIL*, 17, 76-77.

Jha, P. (2006). Digitization and documentation of manuscripts and other cultural heritage at the IGNCA. *Kriti Rakshana*, 2(3), 25-26.

Microsoft India (2005). NIC helps document and digitize Indian manuscripts: Microsoft .NET customer solution case study. Retrieved 15 July 2008 from www.microsoft.com/india

Ministry of Communication and Information Technology (2004). Digitization of culture - background note for Asia IT Ministers' 2nd Summit, Hyderabad. Retrieved 15 June 2008 from www.asiaitsummit.nic.in/Digitisation.pdf

National Mission for Manuscripts (2004). *Guidelines for digitization of manuscripts*. New Delhi: National Mission for Manuscripts.

National Mission for Manuscripts (2005). *Basic minimum standards for conservation of manuscripts.* New Delhi: National Mission for Manuscripts.

National Mission for Manuscripts (2007). *Report of the fourth year: 7 February 2006 – 7 February 2007*. New Delhi: National Mission for Manuscripts.

Paliwal, N. (2007). Vijnananidhi: manuscript treasures of India. *Kriti Rakshana*, 2(4), 6-7.

Paul, S. & Jain, P.K. (2008). Digitization of special libraries in India: with special reference to Indira Gandhi National Centre for Arts (IGNCA) library. Presented in the *ICIKM Conference 2008*, Kathmandu, Nepal.

UNESCO (2001). *Universal Declaration on Cultural Diversity*. Retrieved 15 July 2008 from www.un-documents.net/udcd.htm

UNESCO (2002). *Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives.* Retrieved 15 July 2008 from www.ifla.org/VII/s19/pubs/digit-guide.pdf

CHAPTER 5

Digitization of Indian Journals

5.1 Introduction

With the emergence of information and communication technologies (ICT) in academic and research environment – not only in developed countries but also in developing nations – scholars have become more active and empowered in the processes of scholarly communication and information absorption. Modern ICT-enabled research environment helps the institutions in realization of research goals by electronically communicating results of research in the form of research papers to journal publishers. Similarly, institution-wide access to primary and secondary periodicals is implemented utilizing ICT infrastructure available within the institutions. This enables researchers to get timely information on published literature as well as forthcoming literature just before publishing in print journals. Now-a-days e-journals are more familiar to academic communities than even a decade before. This is due to their far-wider impact on the research communication process as well as to the research communities (Canessa & Zennaro, 2008; European Commission, 2008).

While commercial journal publishers have quite impressive journal access mechanisms due to their early adoption of electronic publishing model, non-commercial journal publishers have started lately to provide full-text access to their journal contents.

To the journal publishers, electronic journal publishing is more cost-effective option than its print counterpart, particularly for journals published by non-profit institutions and societies. In India, most of the primary journals are published by non-profit bodies (Sen, 2002). However, these journals were not properly promoted and merchandised to the international markets when only print versions were available.

With the emergence of knowledge societies, knowledge-based products find wider marketplace. Journal itself is a knowledge-based product and India being an emerging knowledge economy, Indian journals published by the world famous institutions in India are gaining much importance to global audience. As these Indian journals did not have proper marketing and promotional channel abroad, commercial multinational publishers have come forward to tap this potential journal segment. There is also possibility of taking over control of research journals from the scientific societies by the commercial publishers that already happened to academic institutions in developed countries. This may increase the visibility and prestige of journals, but may decrease the availability as compared to open access journals. Corporatization of academic journals can also have impact on editorial autonomy.

The transition from print journal publishing to electronic publishing is not very easy for non-commercial publishers. Profit making journal publishers have some good reasons for investing on digitization of their journal back volumes. One of the most important reasons is that journal publisher can recover its investment within a shorter span of time by providing fee-based subscription to back volumes of the journals. However, the same model cannot be applied to journals published from the developing countries or journals published by the non-profit bodies. Thus, journals published by non-profit bodies have two feasible solutions to this problem: (i) collaboration with commercial journal publishing house that has internationally visible e-journal gateway, or (ii) digitization and web-hosting of back volumes of journals through government-funded projects. The first solution leads to the closed access subscription-based model, while the second solution is the open access model promoted by many public institutions, scientific academies and funding agencies in developed as well as developing countries.

Web-archiving of journal articles can be started from any volume. But majority of the journals prefer to digitize their back issues from volume one, if the digitization fund permits (Ghosh & Das, 2007).

Figure 5.1 indicates that digitization of back volumes of journals lead to four kinds of products, namely:

(i) Open Access Journals – These are web-based journals worldwide freely available to online communities;

(ii) Subscription-base e-Journals - These are web-based closed access fee-based journals worldwide available to the paid online subscribers;

(iii) CD-ROM Collection of Journal Articles - These are CD-ROM-based offline collection of journal articles available mainly to the paid subscribers. CD-ROM collection is very useful to the institutions that have no or limited Internet connectivity;

(iv) Intranet Collection of Journal Articles – This database of journal articles is available and accessible to the end users within a campus boundary having many Intranet nodes for accessing from a central server.

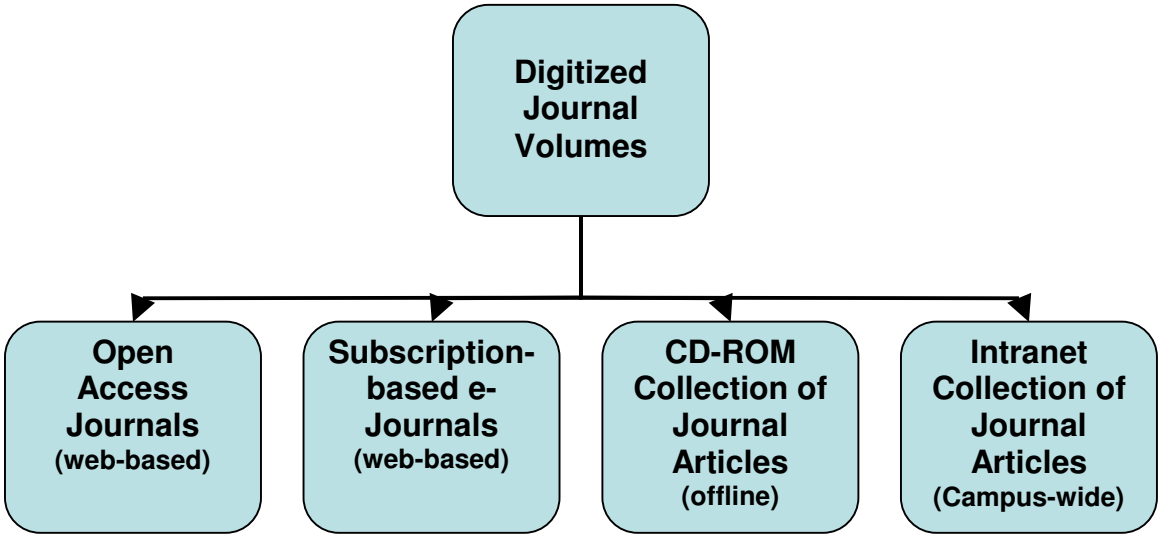


Fig. 5.1: Modes of Access to Digitized Journal Collections

The scholars across the world review not only current literature but also consult old research publications for their academic and research work. As current research is primarily based on research done in past, literature review becomes

an essential process to comprehend the past works. When an online information system provides full-text access to journal literature, it not only covers current contents but also expected to provide archive of past volumes. However, most of the journal publishers in the South Asian region and in other developing countries do not have electronic copies of past articles published a long time ago. They even started electronically processing of manuscripts for submitted or accepted papers very recently (Sahu, 2008; Sahu & Parmar, 2006). Thus, born digital journal articles are very recent phenomena in developing countries including in India. Old articles are available in printed volumes, and publishers are realising importance of making them available online. A retrospective digitization project for each journal aims at converting printed articles into digitized articles that can be archived in digital media. Full-text information storage and retrieval system is also placed to take care of digitized contents of journal articles. Digitization project also converts printed papers into searchable digital texts in order to facilitate full-text searching.

Each online journal gateway is built upon efficient search engine, where many search parameters are available such as article title, name of author, keyword, and subject. Some filtering option is also available in the search interface such as name of journal, publishing year, volume number or subject. Most of the journal gateways use Dublin Core Metadata Standards where metadata elements are embedded in XML[1] files.

Figure 5.2 indicates a simple process of web-enablement of journal articles that involves web-archiving of born digital as well as digitized contents. For both types of contents the journal gateway provides single comprehensive searching and browsing interface. This Figure also signifies that web-archiving of journal contents makes the way to open access journals. A majority of open access journals in India started with journal digitization projects and later convinced to adopt open access model for reaching out to the worldwide audiences.

[1] XML (Extensible Markup Language) is an open standard suitable for web publishing.

Figure 5.3 indicates other two channels of archiving digitized journal articles, i.e., (i) through subject repository and (ii) through institutional repository, in addition to journal gateway. For example, digitized contents of *Indian Journal of Tuberculosis* and some other journals are archived in the popular subject repository in India - OpenMED@NIC. Similarly, the institutional repository of Indian Institute of Astrophysics in Bangalore provides full-text access to digitized articles of *Bulletin of the Astronomical Society of India*.

As indicated earlier, digitized articles can also be part of online full-text research databases as well as publisher-specific online journal gateways. The full-text database service providers aggregate contents from different journal publishers. These contents are not current literature – publishers usually provide one year backward journal contents. On the other hand publisher-specific journal gateway provides access to current as well as back volumes of journal contents. While research database can be available both as CD-ROM collections and online collections, this service is fee-based and subscribers can only opt for bouquet of journals – not individual journals. Similarly, publisher-specific online journal gateway is fee-based information service. However, some publishers provide flexibility in choosing a set of journals, based on institutional requirements of the subscribers.

Figure 5.4 provides a subscription-based diffusion model for digitized journal articles. There are some online secondary databases originated from India that also provide web-links to full-text research contents along with index to journal articles. However, these databases do not exactly provide aggregation service, instead they act as metadata harvesters or indexers for facilitating full-text access to journal articles. Examples of these services are *INDmed* of National Informatics Centre, *Open J-Gate* of Informatics India Limited, and *Open Index Initiative* of Indira Gandhi Institute of Development Research (IGIDR).

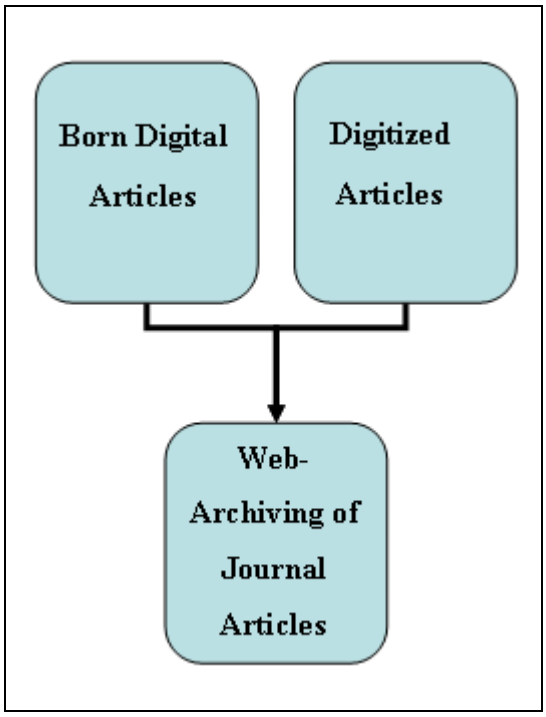


Fig. 5.2: Web-Archiving of Journal Articles – Making the Way to Open Access Journals

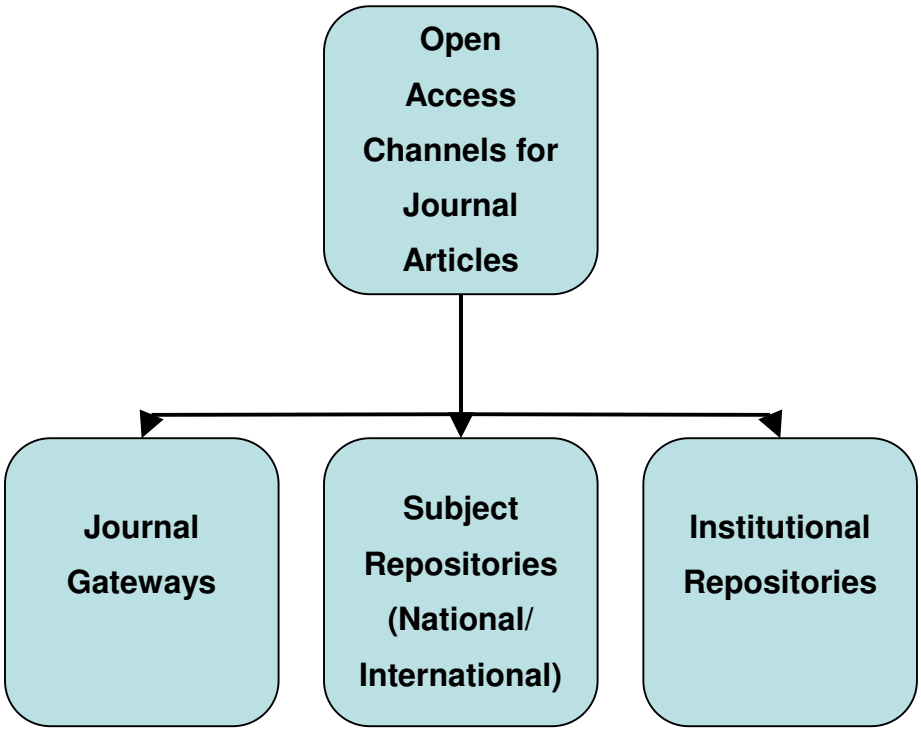


Fig. 5.3: Open Access Channels for Digitized Journal Articles

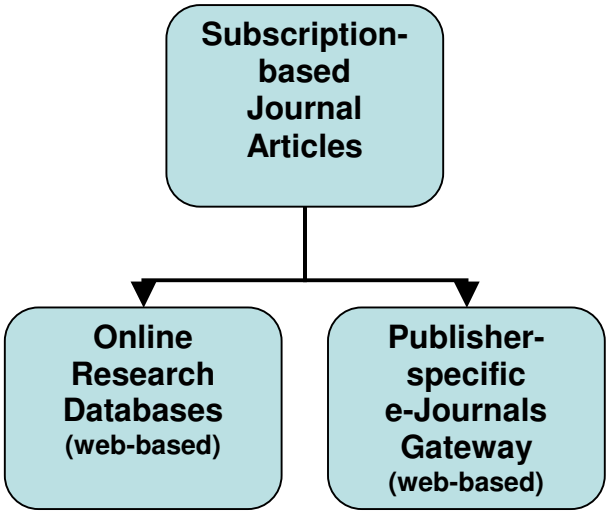


Fig. 5.4: Subscription-based Access to Digitized Journal Articles

5.2 Transformation of Indian Journal Publications

India has a long tradition of publishing scholarly journals, mostly by the non-profit making bodies such as scientific academies, professional societies, academic institutions, research institutions and government departments. Indian Journals are published in English language as well as in other major Indian languages. The Department of Science and Technology (DST) of Ministry of Science and Technology provides grant-in-aid to five scientific academies in India. Most of these academies publish peer-reviewed scientific periodicals for the wide dissemination of research results, contributed by Indian scientists including fellows of these academies. Figure 5.5 indicates a list of grant-in-aid scientific academies in India supported by DST. Table 5.1 provides journal publishing profile of these scientific academies. This Table also indicates that only two academies, namely Indian National Science Academy and Indian Academy of Sciences, made journal publishing as a thrust area for them. They publish 15 scientific periodicals which are now available both in print and electronic formats.

Scholarly journals of different areas of knowledge are published by Indian institutions and professional societies with support from government bodies and parent bodies. Different government departments have some schemes to support publishing and e-publishing of scholarly journals. In each year DST

provides support for publication of about twenty five important journals published by scientific societies in India through its ongoing scheme called “Assistance to Professional Bodies”. Similarly, Department of Scientific and Industrial Research (DSIR) of Ministry of Science and Technology provides support for e-publishing of existing scientific periodicals through its ongoing scheme called “Electronic Publishing of Selected Indian S&T materials”. Indian research councils such as Council of Scientific and Industrial Research (CSIR) and Indian Council of Medical Research (ICMR) also have mandates to support publications of Indian research periodicals.

With the realization of importance of electronic journals in scholarly communication processes, specially advocated by International bodies such ICSU (International Council of Science), INASP (International Network for the Availability of Scientific Publications) and UNESCO, two Indian scientific academies namely - Indian National Science Academy (INSA) and Indian Academy of Sciences (IAS) – joined together in late 1990s and formed the Joint Academy’s Committee for Electronic Publishing (JACEP) (Mujoo-Munshi, 2003). The JACEP undertook a review of electronic publishing models as followed by scientific societies/academies in developing countries as well as in developed countries. After consultation of editorial committee members of academy journals, fellows of both academies and other stakeholders, the JACEP recommended:

- Digitization of back-volumes of all journals published by INSA and IAS;
- Web-archiving of back-volumes of all journals published by INSA and IAS;
- Open Access to all journals published by INSA and IAS; and
- Each Academy is to set up a web server for hosting journal contents online.

JACEP also recommended an action plan for achieving electronic publishing goals of these two academies. The action plan gave much importance on the capacity development and capacity building of the academies. The ICT

infrastructure of the academies was considered as the capacity development exercise and staff training as the capacity building exercise involving the Supercomputer Education and Research Center (SERC) at the Indian Institute of Science (IISc), Bangalore. Figure 5.6 showing a schematic structure of the JACEP, indicating participating institutions in this effort along with a broad list of recommendations of the JACEP.

Following the JACEP recommendations, INSA and IAS both initiated digitization of back volumes of their journals and later make them globally freely accessible through web portals. This was the beginning of electronic publishing of Indian periodicals that triggered other Indian institutions and journal publishers adopting similar model for worldwide dissemination of research literature originated mainly from India and other south Asian countries.

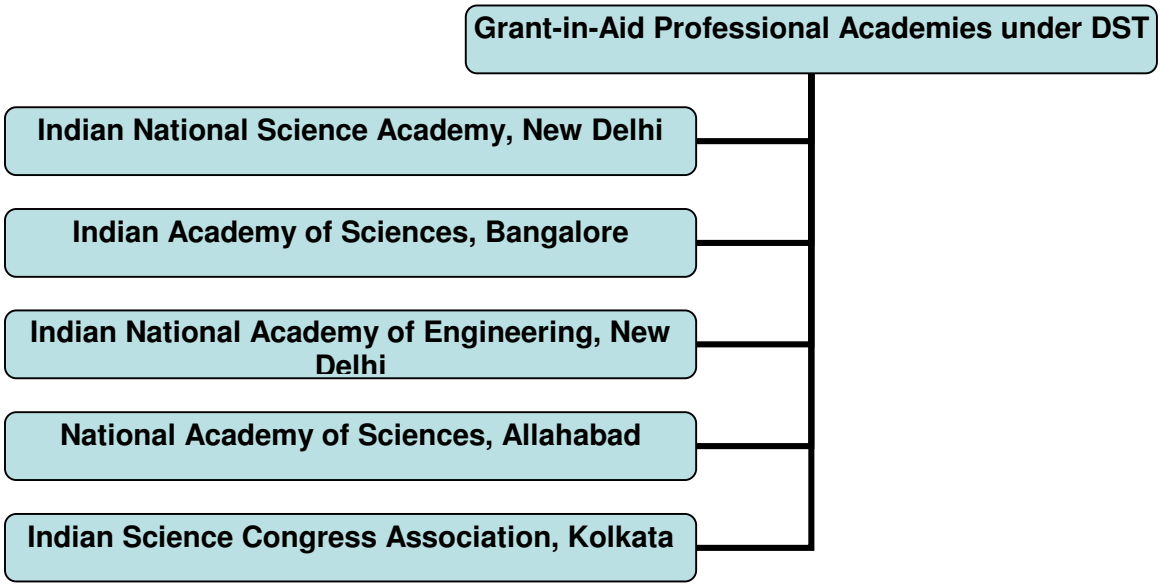


Fig. 5.5: Grant-in-Aid Science Academies in India supported by DST

Table 5.1: Scientific Periodicals Published by the Science Academies in India

	Number of Print Journals	Number of e-Journals
Indian National Science Academy, New Delhi	4	4
Indian Academy of Sciences, Bangalore	11	11
Indian National Academy of Engineering, New Delhi	Nil	Nil
National Academy of Sciences, Allahabad	2	Nil
Indian Science Congress Association, Kolkata	1	1

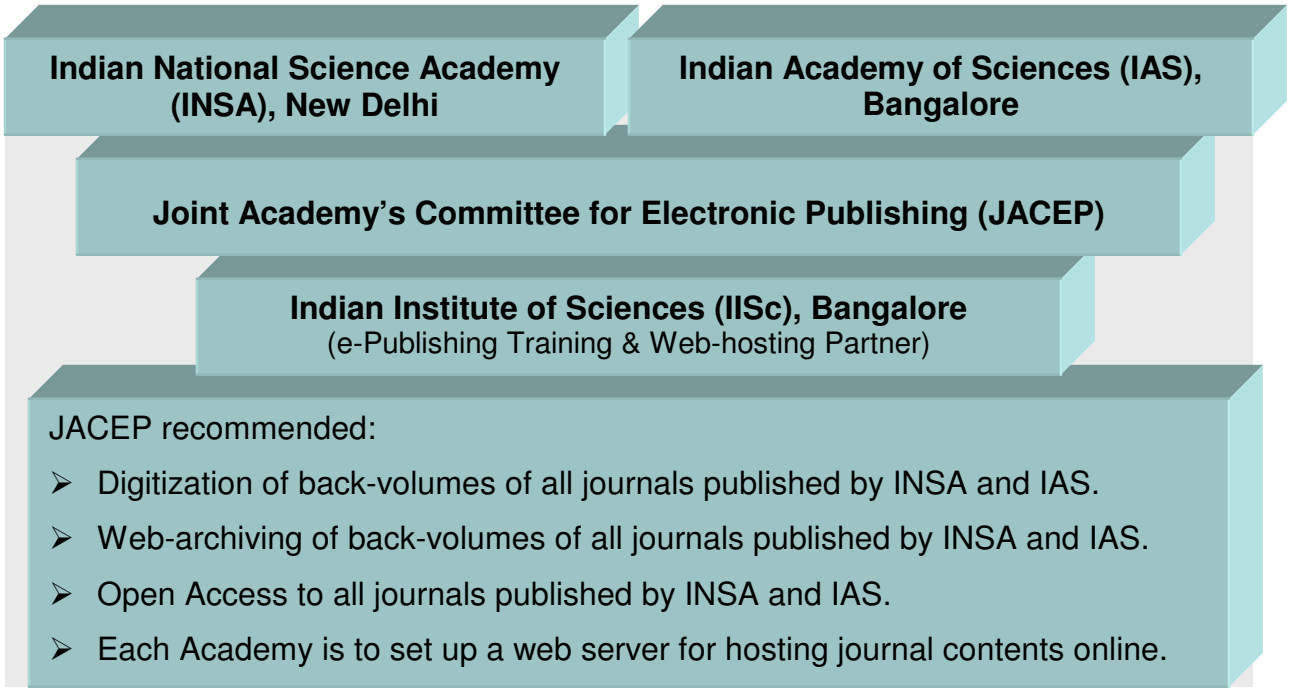


Fig. 5.6: Schematic Structure of JACEP, a Joint Effort of INSA and IAS

5.3 Indian National Science Academy (INSA) [www.insa.ac.in]

The Indian National Science Academy (INSA) is a scientific academy funded by the Government of India. It was established in 1935 as the National Institute of Sciences of India. INSA is a national level professional body of Indian scientists representing all disciplines of science and technology. Its board objectives comprise promotion of science in India, S&T policymaking at the national and state level, safeguarding the interests of the scientists, establishing international linkages with collaborations and partnerships, and exchanging knowledge for the betterment of the society. INSA is engaged in advocating open access concept to Indian policymakers and scholarly publishers (Mujoo-Munshi, 2003).

INSA publishes four peer-reviewed journals, organizes scientific discussions and brings out proceedings and monographs. INSA started publishing research journal as early as in 1935 – in the same year of its establishment. Its first journal was *Proceedings of the National Institute of Sciences of India* that later became *Proceedings of the Indian National Science Academy Part A & B*, as indicated in Table 5.3.

The e-Journals@INSA is an electronic publishing initiative of the Academy that was initiated in July 2002. This project initially started from a NISSAT-supported project titled "Building Digital Resources: Creating Facilities at INSA for hosting S&T Journals on Online" to facilitate conversion of INSA journals from print to digital format and host these materials online. As indicated in Figure 5.7, initial objectives of the NISSAT-supported project were:

- To build up a national digital resource base with global visibility and accessibility;
- To bring out electronic versions of INSA journals and to make them available on a web server; and
- To support and research activities by providing all time access to the full-text scientific information globally.

In later stage, few other project objectives emerged to make this project more invasive and inclusive.

Figure 5.8 shows a schematic structure of the e-Journals@INSA project. This Figure indicates that Digital Publishing Solutions (DPS) in Pune was the operational partner for this project. INSA outsourced the digitization work of this project to the private agency DPS. As indicated in Figure 5.9, DPS was engaged in all round digitization solutions such as digitization of back volumes of INSA journals, creation of metadata for each journal article, development of web-based full-text information storage and retrieval system including search interface and search result display interfaces, and finally web-enablement of the digitized contents along with metadata element sets.

The e-Journals@INSA portal provides access to current and back volumes of full-text literature of INSA journals. Table 5.2 provides a list of current journal contents available in this portal. Table 5.3 provides a list of old titles of INSA journals digitized in this project. Tables 5.2 and 5.3 also indicate that retrospective contents digitized in this project from the volume one of each journal till the year 2002. From 2003 onwards all articles archived in this web-based system are born digital and made available immediately after publishing the journal issues.

All INSA journals are now open access and full-text is available in PDF files from the common journal gateway. This gateway is freely accessible if the users register their names in the website. This portal provides a search interface for advanced search as shown in Figure 5.10, where search query can be given with any or more than one metadata elements, such as keyword, article title, name of the author, journal volume-issue-initial page number, and journal year. The combination of search elements is made possible using Boolean operators. The recent journal issues as well as back volumes of INSA journals can be browsed from the main page of this portal. The search result displays a list of articles that match the search query; when a user selects an article from the search result it leads to the particular journal issue that published the particular reference, the PDF version of article appears on the screen and can be downloaded into the user workstation. Table 5.4 provides a list of metadata elements used for

navigation of the portal, searching the full-text database, and displaying the retrieved documents after successful search queries.

Although this project is one of first open access journal gateways in India that promoted the concept of free access to scientific literature, this gateway is not compliant to the Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH) - the international standard for web content interoperability. Thus, metadata harvesting and indexing service providers for open access journals such as Google Scholar, Directory of Open Access Journals (DOAJ) and Open J-Gate, cannot index INSA journals archived in this gateway as indicated in Table 5.5. Worldwide scientists and researchers prefer to use secondary services such as metadata harvesting and indexing services for open access journals to identify relevant literature in their areas of interest. Thus, INSA loses crucial international visibility of their journals by not adopting OAI-PMH-compliant journal publishing model.

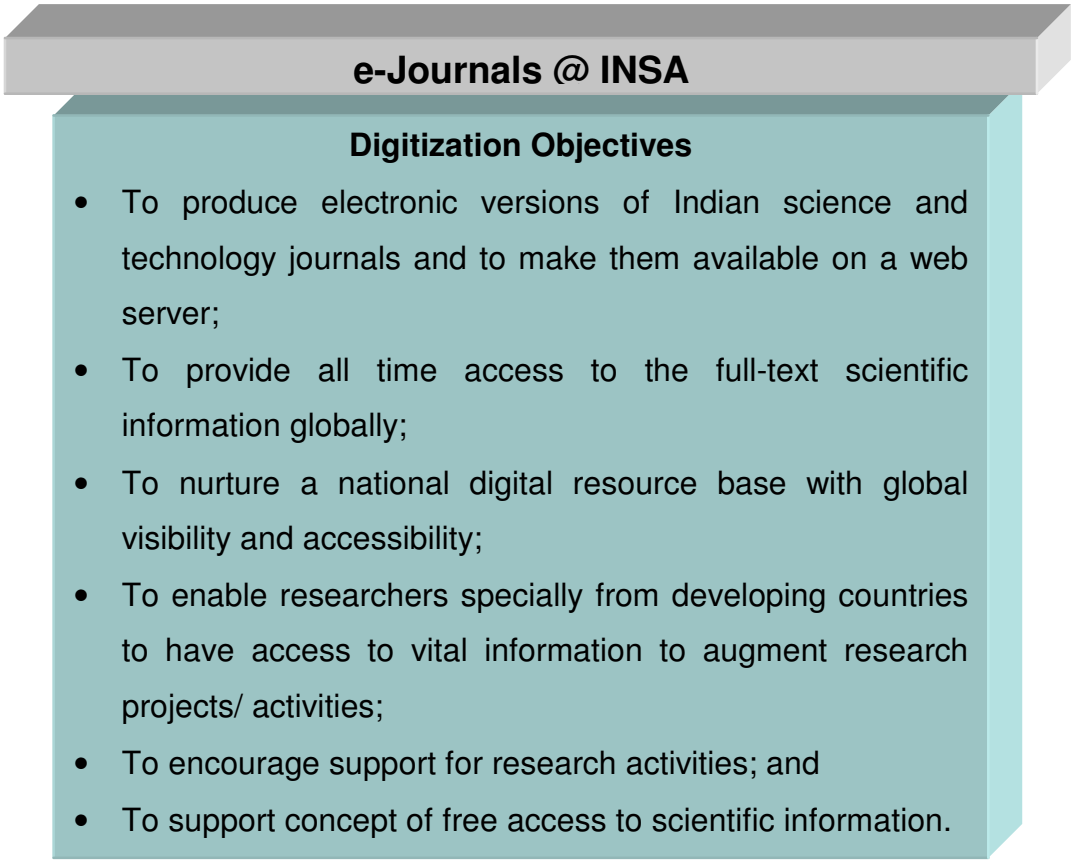


Fig. 5.7: Broad Objectives of e-Journals @ INSA Project

NISSAT = National Information System for Science and Technology
DSIR = Department of Scientific and Industrial Research
GoI = Government of India

Fig. 5.8: Schematic Structure of e-Journals @ INSA Project

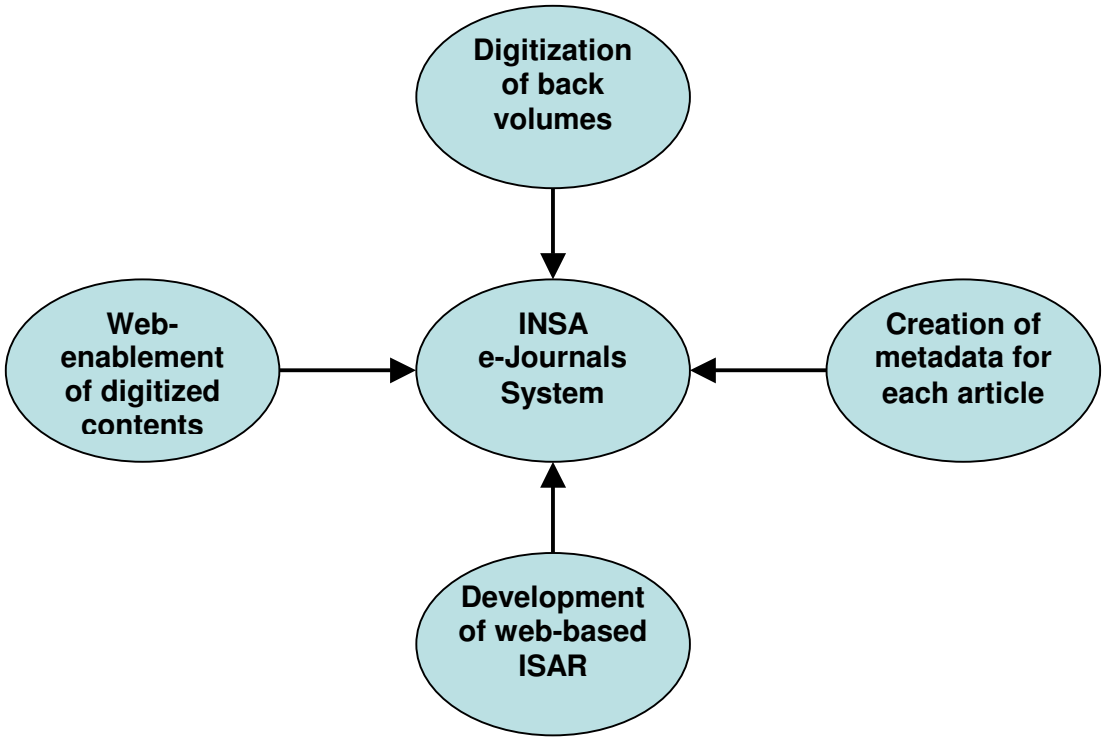


Fig. 5.9: Integration of different applications into INSA e-Journals System

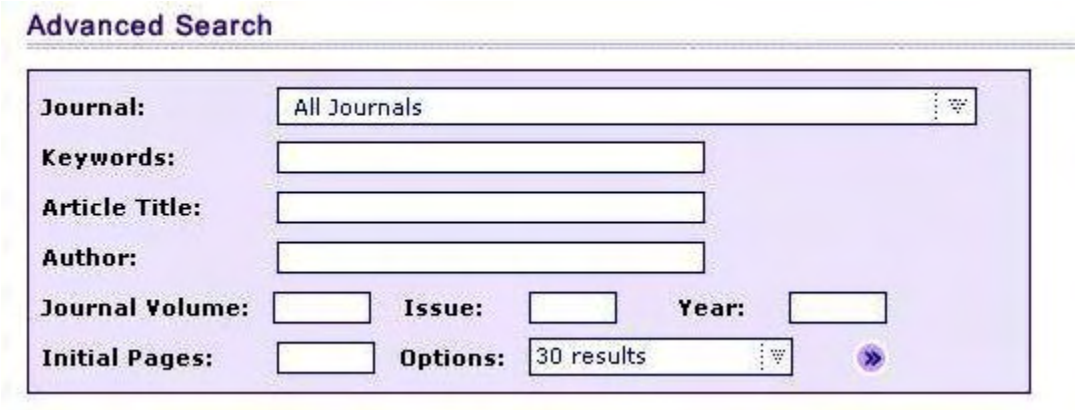


Fig. 5.10: Advanced Search Interface in INSA e-Journals System

Table 5.2: Digitization of INSA Journals – Current Titles

Journal Name	No. of issues a year	Full-Text Availability from Year	Contents Digitized	Born Digital Contents
Indian Journal of Pure and Applied Mathematics	12	Since 1970	1970-2002	Since 2003
Indian Journal of History of Science	4	Since 1966	1966-2002	Since 2003
Proceedings of the Indian National Science Academy Part A - Physical Sciences	6	Since 1970	1970-2002	Since 2003
Proceedings of the Indian National Science Academy Part B - Biological Sciences	6	Since 1970	1970-2002	Since 2003

Table 5.3: Digitization of INSA Journals – Old Titles

Journal Name	Full-Text Contents Digitized
Proceedings of the National Institute of Sciences of India (Bifurcated into two parts in 1955 and published as Proceedings of National Institute of Sciences of India - Part A and Part B)	1935-1954
Proceedings of the National Institute of Sciences of India Part A - Physical Sciences (Re-christened as Proceedings of the Indian National Science Academy Part A & B)	1955-69
Proceedings of the National Institute of Sciences of India Part B - Biological Sciences (Re-christened as Proceedings of the Indian National Science Academy Part A & B)	1955-69

Table 5.4: Metadata Elements Used in e-Journals@INSA

Browsable Metadata	Searchable Metadata	Displayed Metadata
➢ Current Journal Issue ➢ Journal Back Issues (Name of Journal, Volume, Year, and Issue Number)	➢ Journal Name ➢ Keywords ➢ Article Title ➢ Author ➢ Journal Volume ➢ Issue ➢ Year ➢ Initial Pages	➢ Article Title ➢ Author ➢ Keywords ➢ Initial Page ➢ Journal Name ➢ Volume ➢ Issue ➢ Year ➢ Read Article (Hyperlink to Full-text) ➢ More Articles in This Issue (Hyperlink)

Table 5.5: Coverage of INSA Journals by Different Online Gateways

Journal Name	DOAJ	Google Scholar	Open J-Gate
Indian Journal of Pure and Applied Mathematics	X	X	√
Indian Journal of History of Science	X	X	X
Proceedings of the Indian National Science Academy Part A - Physical Sciences	X	X	X
Proceedings of the Indian National Science Academy Part B - Biological Sciences	X	X	X
Coverage	0%	0%	25%

5.4 Indian Academy of Sciences [www.ias.ac.in/pubs/journals/]

The Indian Academy of Sciences (IAS) is an important scientific academy in India that helps in strengthening research in the basic and applied science areas, promoting scientific temperament and improving science education in the country. The Academy was established and registered as a scientific society in 1934. IAS is involved in organizing meetings and discussions on important topics, recognizing scientific talents through its Fellowship programme, improvement of science education and taking up other issues of concern to the scientific community. Presently focused action areas of IAS include science education, women in science and scientific data of public interest. Academy attempts to fulfil its broad objectives through original research and dissemination of scientific knowledge to the community via its meetings, discussions, seminars, symposia and publications. IAS is the biggest academy in the country publishing peer-reviewed journals covering all major scientific disciplines. IAS journals publish papers submitted by its fellows as well as non-fellow researchers from different countries. IAS began its publishing journey in 1934 with Academy Proceedings that bifurcated into many parts in later time to keep up with the emerging areas of scientific research.

The Academy collaborates with different institutions in the country and even abroad for publishing a wide range of scientific journals. IAS receives grant-in-aid

from the Department of Science Technology (DST) of Ministry of Science and Technology (MST), Government of India. Due to DST support in carrying out its electronic publishing activities, all journals of IAS are now 'open access' and full text is available as PDF files on each journal's website.

Figure 5.11 indicates a collaboration pattern in bringing out Academy journals and dissemination of scientific literature through e-publishing gateways. In association with Current Science Association and Indian Institute of Sciences, Academy publishes *Current Science*, which is a widely circulated fortnightly journal on S&T progress in India. IAS also publishes the *Journal of Genetics*, the oldest journal of its kind in the English language that started in 1910. This journal was on the verge of closure and in 1985 the Academy took over and revived this journal. Other than publishing research journals, in 1996 IAS added a popular science journal entitled Resonance that aims at improving the quality of science education and teaching.

In December 2006, the IAS entered into an agreement with one of the leading international publishers Springer for co-publishing ten journals of the IAS. This collaboration is aimed enhancing international visibility of all selected journals published by IAS. Springer helps in improving the overseas subscriptions and also makes available all the Academy journals on their website with enhanced search and retrieval features. The journals also continue to be freely available online from the server of the Academy.

This co-publishing initiative by a global publishing leader recognizes that India has become one of the top-ranking countries in the field of basic science research and Indian Science has become the most powerful instruments for growth and development of the country.

The IAS publications portal provides access to current and back volumes of full-text literature of Academy journals. Table 5.6 provides a list of journal contents available in this portal. This Table also indicates frequency of each journal and individual journal’s site address.

All of the articles in current issues of Academy journals are born-digital from 2000 onwards. The articles of back volumes, which were not born-digital, were digitized through a government-supported project. IAS has already archived all articles of eight journals from the first volume whereas full-text contents of other three journals are available from some old volumes. Born digital current journal issues are made available immediately after publishing the journal issues. Table 5.7 indicates full-text availability of Academy journals in IAS journal gateway and SpringerLink portal .

The Academy publications portal provides easy navigation facility to the users. The current journal issue as well as back volume of a journal can be browsed from the main page of each journal's website. The user has to first select year or volume of a journal then proceed to navigate table of contents of an issue of the selected volume. When user selects an article from the table of contents, the PDF version of that article appears on the screen and can be downloaded into the user workstation.

Unlike INSA journals, IAS journals are compliant to the Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH). Many indexing service providers for open access journals effectively index and harvest contents of almost all IAS journals archived in its portal. Table 5.8 shows a matrix of coverage of IAS journals by online journal indexing and metadata harvesting services such as Google Scholar, Directory of Open Access Journals (DOAJ) and Open J-Gate. This Table also indicates availability of IAS journals in SpringerLink gateway that has very efficient searching and retrieval interface.

Open access to IAS journals increases their global visibility and worldwide acceptance. Internationally acclaimed online research databases such as Web of Knowledge and Scopus have considered this fact and now index most of the IAS journals. Indian scientific periodicals are being considered in these citation databases due to worldwide recognition of publications' quality maintained by the Academy.

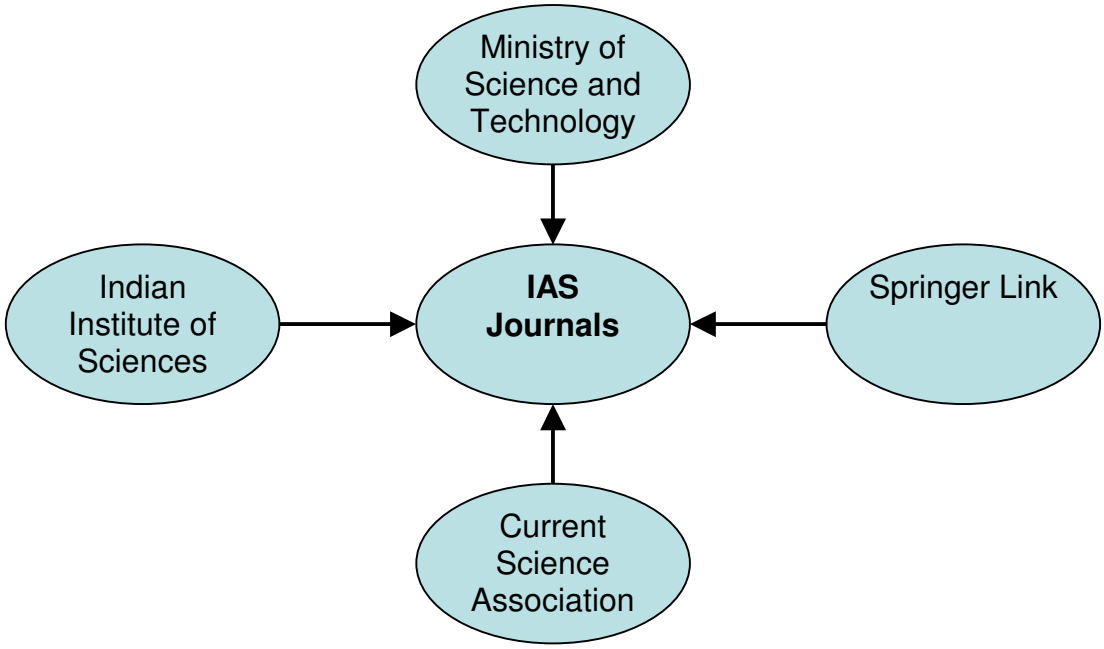


Fig. 5.11: Partnership Pattern for Electronic Publishing of IAS Journals

Table 5.6: Indian Academy of Sciences Published Open Access Journals

Journal Name	No. of issues a year	Web Address of Journal Portal
Current Science	24	www.ias.ac.in/currsci/
Journal of Chemical Sciences	6	www.ias.ac.in/chemsci/
Proceedings of the Indian Academy of Sciences: Mathematical Sciences	4	www.ias.ac.in/mathsci/
Journal of Earth System Science	4	www.ias.ac.in/jess/
Sadhana (Academy Proceedings in Engineering Sciences)	6	www.ias.ac.in/sadhana/
Pramana - Journal of Physics	12	www.ias.ac.in/pramana/
Journal of Biosciences	4	www.ias.ac.in/jbiosci/
Bulletin of Materials Science	6	www.ias.ac.in/matersci/
Journal of Astrophysics and Astronomy	4	www.ias.ac.in/jaa/
Journal of Genetics	3	www.ias.ac.in/jgenet/
Resonance - Journal of Science Education	12	www.ias.ac.in/resonance/

Table 5.7: Full-text Availability of IAS Journals in Two Journal Gateways

Journal Name	IAS e-Journal Portal		Springer Link	
	Vol.	Year	Vol.	Year
Current Science	1	1932	Nil	Nil
Journal of Chemical Sciences	86	1977	88	1979
Proceedings of the Indian Academy of Sciences: Mathematical Sciences	87	1978	87	1978
Journal of Earth System Science	87	1978	87	1978
Sadhana (Academy Proceedings in Engineering Sciences)	1	1978	1	1978
Pramana - Journal of Physics	1	1973	1	1973
Journal of Biosciences	1	1979	1	1979
Bulletin of Materials Science	1	1979	1	1979
Journal of Astrophysics and Astronomy	1	1980	1	1980
Journal of Genetics	1	1910	64	1985
Resonance - Journal of Science Education	1	1996	1	1996

Table 5.8: Coverage of IAS Journals in OA Journal Indexing Gateways

Journal Name	DOAJ	Google Scholar	Open J-Gate	Springer Link
Current Science	√	√	√	X
Journal of Chemical Sciences	√	√	√	√
Proceedings of the Indian Academy of Sciences: Mathematical Sciences	√	√	√	√
Journal of Earth System Science	√	√	√	√
Sadhana (Academy Proceedings in Engineering Sciences)	√	√	√	√
Pramana - Journal of Physics	√	√	√	√

Journal of Biosciences	√	√	√	√
Bulletin of Materials Science	√	√	√	√
Journal of Astrophysics and Astronomy	√	√	X	√
Journal of Genetics	√	√	√	√
Resonance - Journal of Science Education	X	X	√	√
Total	***10***	***10***	***10***	***10***
Coverage	***90.9%***	***90.9%***	***90.9%***	***90.9%***

5.5 medIND@NIC: Biomedical Journals from India

The Indian MEDLARS Centre is one of the international MEDLARS centres established in different countries. Indian MEDLARS Centre is an advanced medical informatics centre responsible for South Asian region, functioning under the Bibliographic Services Division (BSD) at the National Informatics Centre (NIC) in New Delhi. This unit is engaged in indexing and abstracting of Indian biomedical periodicals for inclusion in the PubMed databases maintained by the United States National Library of Medicine and the National Institutes of Health. PubMed databases include resources from the MEDLARS (Medical Literature Analysis and Retrieval System) or MEDLINE databases.

As indicated in Figure 5.12, Bibliographic Services Division (BSD) has a wide range of information products – two full-text databases and three reference databases (Singh, Gaba & Pandita, 2004). These information products are designed, developed and web-enabled at in-house facilities in NIC. A single window national online bibliographic database IndMED@NIC indexes 70+ prominent biomedical journals of India from 1985 onwards. While IndMED is an online reference database accessible worldwide, another project of NIC 'medIND@NIC' supplements IndMED@NIC service by providing open access to the full-text contents of 39 Indian biomedical journals covered in IndMED@NIC

database. This Figure also shows other full-text and reference services maintained by Bibliographic Services Division of NIC.

Figure 5.13 shows a schematic structure of medIND@NIC project. This Figure indicates that Indian Council of Medical Research (ICMR) has functional collaboration in terms of financial support, while Indian MEDLARS Centre of Bibliographic Services Division at NIC implements this open access initiative.

medIND aims at providing online access to full-text Indian biomedical periodicals to the users within and outside India. Different publishers, mainly learned societies in the respective specialized areas, publish medIND covered journals in print-on-paper format. Figure 5.14 shows the linkages between publishers, ICMR and NIC in the process of implementation of this project. Figure 5.15 identifies different categories of publishers that provide full-text journal contents to this initiative. The content providers are drawn from professional societies, medical schools, R&D institutions and a medical council ICMR.

Figure 5.16 shows a digitization workflow in medIND project that starts with identifying journals for digitization. IndMED covered journals are only considered for this digitization project as these journals maintain high quality, usually publish in time and have fewer backlogs. After identifying journals, publishers of these journals are invited to establish an official agreement through memorandum of understanding (MOU). The publishers who signed up legal agreements send their journal back volumes to NIC for digitization. NIC digitizes the contents of back volumes of these journals. NIC also establishes linkages of digitized articles with IndMED database, if the bibliographic records are already available there. Otherwise, new bibliographic record for each digitized article is created in IndMED database. Finally NIC web-enables full-text contents of digitized journals to make them freely accessible through medIND open access journal gateway. Initially, medIND project digitized some back volumes of journals. But now NIC is mostly receiving born-digital contents of current issues from the publishers. Some of these publishers also maintain separate journal websites for providing full-text access to journal contents.

Table 5.9 provides full list of Indian journals archived in medIND gateway. This Table also shows a status of the availability of full-text contents of different journals. This Table also indicates that in the medIND project only fewer volumes were digitized mostly after year 2000 period until the born digital contents were made available.

Table 5.10 provides names of publishers along with respective journal titles that belong to different categories of journal publishers as indicated in Figure 5.15. A close observation of names of publishers suggests that a majority of journal publishers belongs to the professional society category and other publishers belong to institutional categories. Figure 5.17 shows a distribution pattern of different kinds of journals in medIND project. This Figure also indicates that professional societies contribute 76.9% of total number of journals whereas professional institutions including biomedical schools, R&D institutions and medical council, contribute 23.1% journals to the medIND gateway.

Similar to IAS journals, medIND journals are compliant to the Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH). Many indexing service providers for open access journals effectively index and harvest contents of almost all medIND journals archived in its portal. Table 5.11 shows a matrix of coverage of medIND journals by online journal indexing and metadata harvesting services such as Google Scholar, Directory of Open Access Journals (DOAJ) and Open J-Gate. This Table also indicates availability of medIND journals in MedKnow gateway that is a private sector professional initiative to make Indian biomedical periodicals internationally visible through open access and widely indexed in secondary research databases. Table 5.11 also indicates that about 35.90% medIND journals are listed in DOAJ, about 25.64% medIND journals are indexed by DOAJ Content service. Indexing services of Google Scholar and Open J-Gate have better coverage for medIND journals with 84.62% and 79.49% coverage respectively.

Table 5.11 and Figure 5.18 show that MedKnow covers about 30.8% medIND journals. While medIND offers merely web-archiving facility to Indian biomedical

journals, MedKnow offers total publishing and e-publishing solutions to biomedical journal publishers in addition to web-archiving service. Thus, MedKnow is gaining popularity amongst Indian biomedical institutions and professional societies, whereas medIND receives cold shoulder from the same professional bodies due to lack of innovative practices in present state of medIND project.

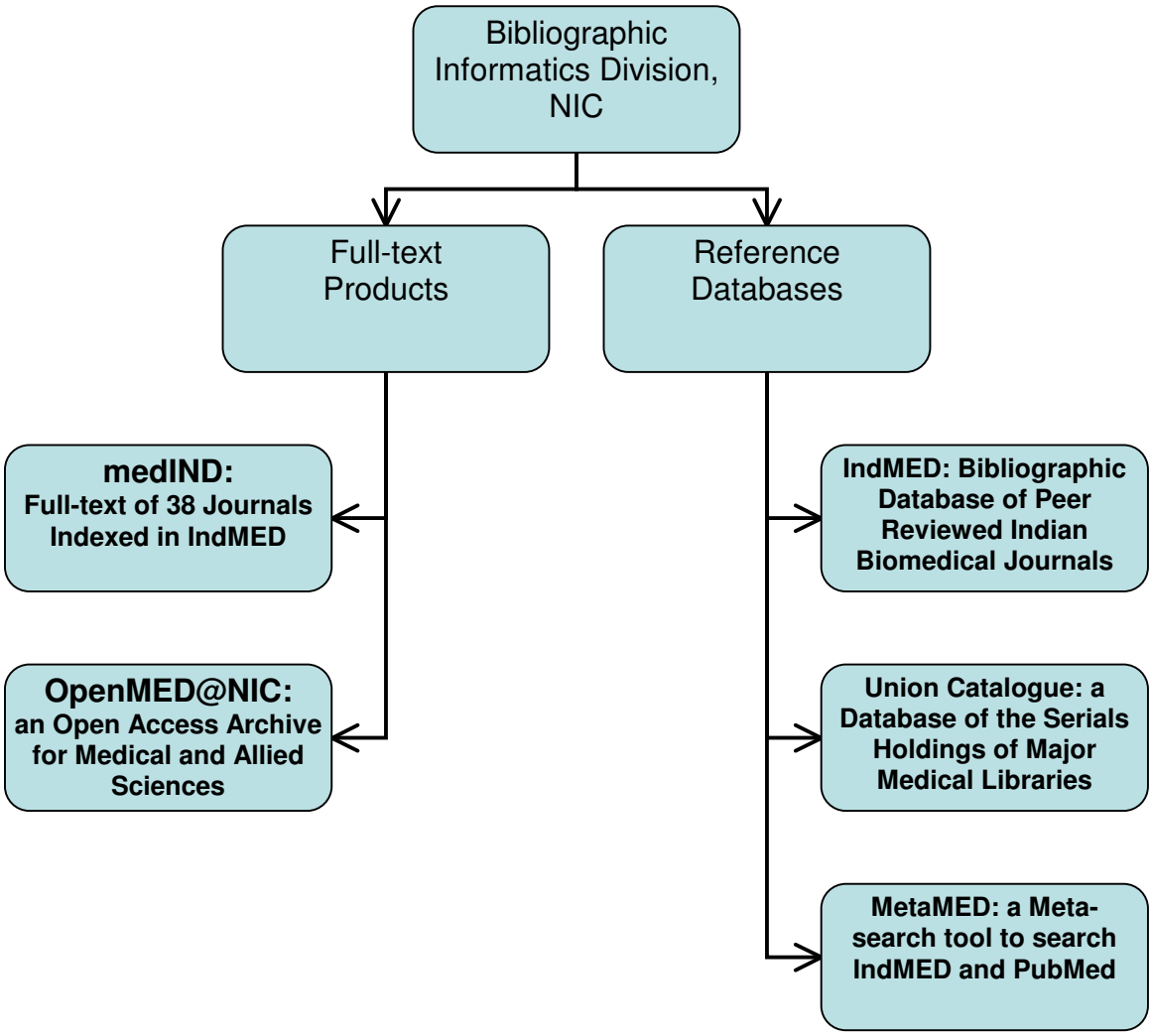


Fig. 5.12: Full-text and other Information Products of Bibliographic Informatics Division at NIC

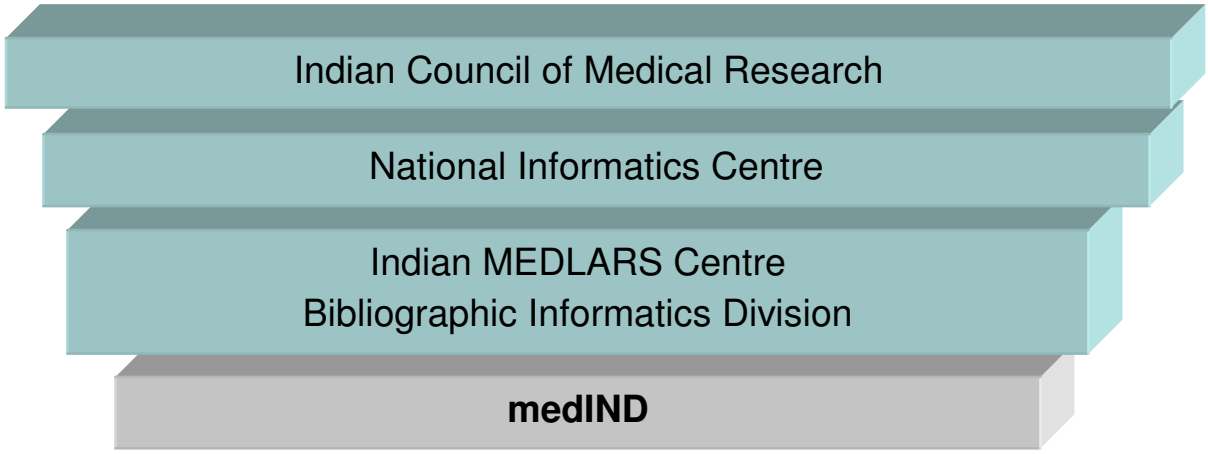


Fig. 5.13: Schematic Structure of medIND Project

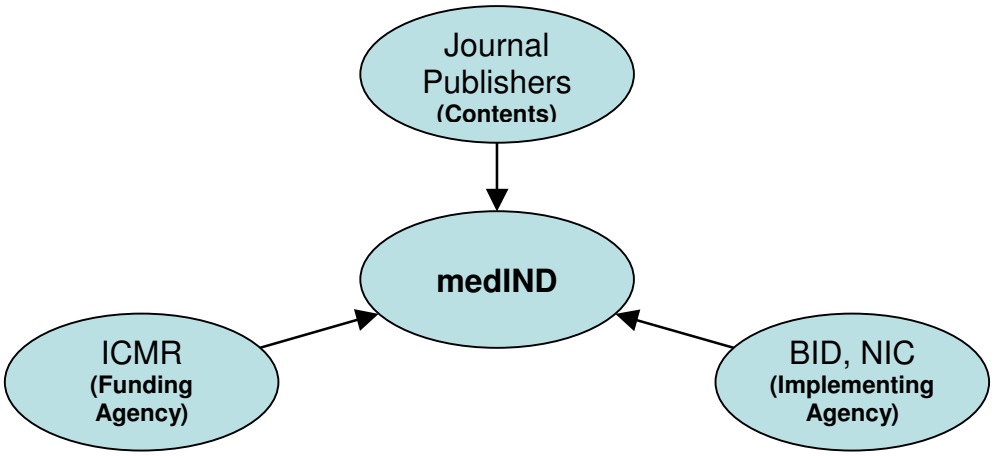


Fig. 5.14: Functional Collaborative in medIND Project

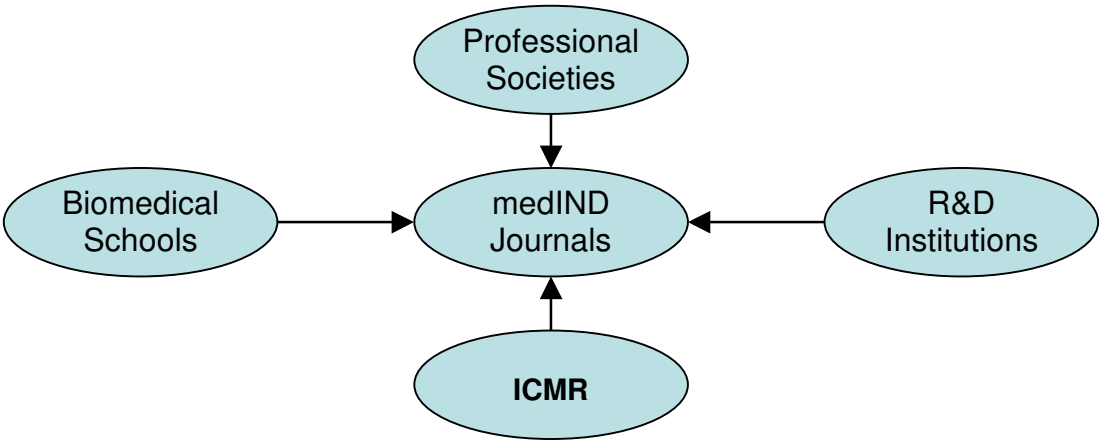


Fig. 5.15: Participation of Different Types of Publishers in medIND Project

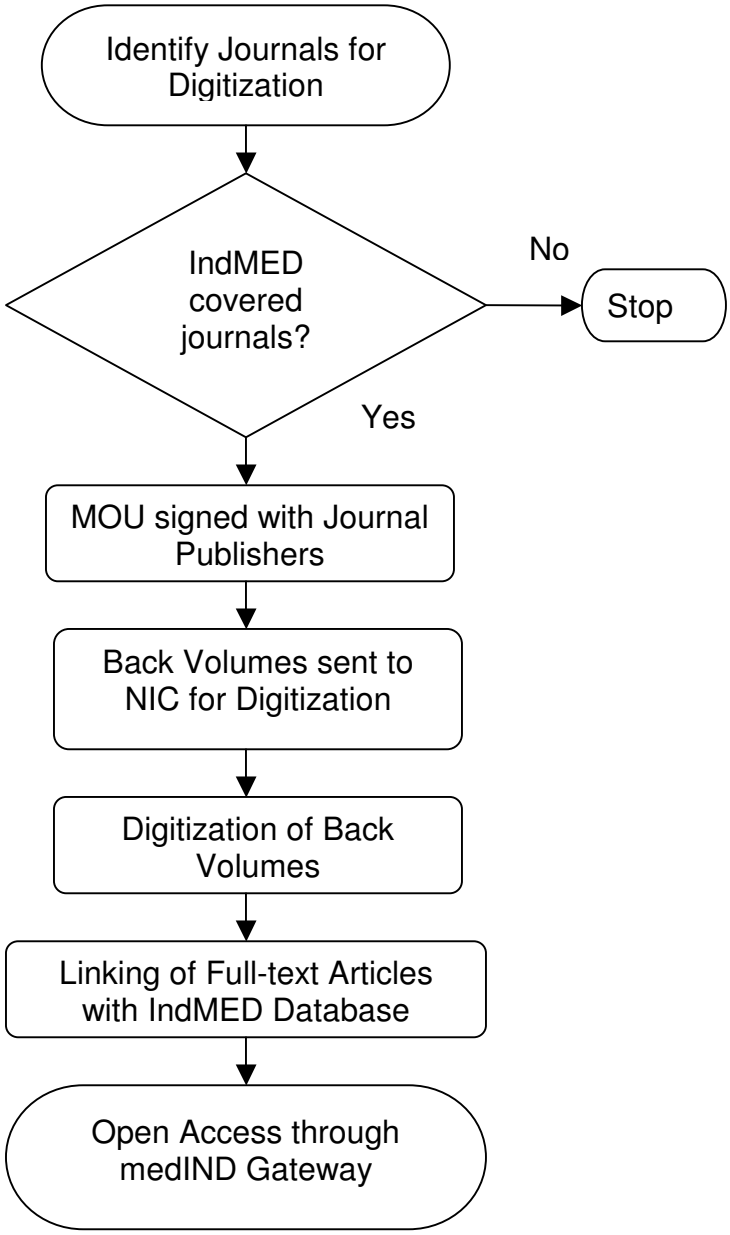


Fig. 5.16: Digitization Workflow in medIND Project

Table 5.9: medIND Hosted Open Access Journals

Journal Name	Full-text Availability	
	Year	Vol. No.
Annals of Cardiac Anaesthesia	2005	8
Endodontology	2000	12
Health Administrator	2000	9
Indian Journal of Aerospace Medicine	2000	44
Indian Journal of Allergy Asthma and Immunology	2000	14
Indian Journal of Anaesthesia	2002	46
Indian Journal of Chest Diseases and Allied Sciences	2000	42
Indian Journal of Clinical Biochemistry	2002	17
Indian Journal of Community Medicine	2000	25
Indian Journal of Gastroenterology	2004	23
Indian Journal of Medical and Paediatric Oncology	2001	22
Indian Journal of Medical Microbiology	2001	19
Indian Journal of Medical Research	2004	119
Indian Journal of Nephrology	2001	11
Indian Journal of Nuclear Medicine	2002	17
Indian Journal of Occupational and Environmental Medicine	2003	7
Indian Journal of Occupational Therapy	2002	33
Indian Journal of Otolaryngology and Head and Neck Surgery	2001	53
Indian Journal of Pediatrics	2005	72
Indian Journal of Pharmacology	2000	32
Indian Journal of Preventive and Social Medicine	2003	34
Indian Journal of Radiology and Imaging	2004	14
Indian Journal of Sexually Transmitted Diseases	2004	25
Indian Journal of Thoracic and Cardiovascular Surgery	2003	19
		Cont...

Cont...		
Journal Name	**Full-text Availability**	
	Year	**Vol. No.**
Indian Journal of Thoracic and Cardiovascular Surgery	2003	19
Indian Journal of Tuberculosis	2000	47
Indian Pediatrics	2004	41
J.K. Practitioner	2001	8
Journal, Indian Academy of Clinical Medicine	2000	5
Journal of Family Welfare	2000	46
Journal of Indian Academy of Forensic Medicine	2004	26
Journal of Indian Rheumatology Association	2002	10
Journal of Obstetrics and Gynecology of India	2004	54
Journal of the Anatomical Society of India	2001	50
Journal of Indian Association of Pediatric Surgeons	2001	6
Journal of Indian Society of Pedodontics and Preventive Dentistry	2000	18
Lung India	2004	21
Medical Journal Armed Forces India	2000	56
NTI Bulletin	2000	36
Trends in Biomaterials and Artificial Organs	2001	15

Table 5.10: Electronic Journal Publishing Partners in medIND Project

Journal Name	Journal Publisher
Annals of Cardiac Anaesthesia	Indian Association of Cardiovascular Thoracic Anaesthesiologists
Endodontology	Indian Endodontic Society
Health Administrator	Indian Society of Health Administrators
Indian Journal of Aerospace Medicine	Indian Society of Aerospace Medicine
Indian Journal of Allergy Asthma and Immunology	Indian College of Allergy Asthma and Immunology
Indian Journal of Anaesthesia	Indian Society of Anaesthesiologists
Indian Journal of Chest Diseases and Allied Sciences	Vallabhbhai Patel Chest Institute, & National College of Chest Physicians
Indian Journal of Clinical Biochemistry	Association of Clinical Biochemists of India
Indian Journal of Community Medicine	Indian Association of Preventive & Social Medicine
Indian Journal of Gastroenterology	Indian Society of Gastroenterology
Indian Journal of Medical and Paediatric Oncology	Indian Society of Medical & Paediatric Oncology
Indian Journal of Medical Microbiology	Indian Association of Medical Microbiologists
Indian Journal of Medical Research	Indian Council of Medical Research
Indian Journal of Nephrology	Indian Society of Nephrology
Indian Journal of Nuclear Medicine	Society of Nuclear Medicine, India
Indian Journal of Occupational and Environmental Medicine	Indian Association of Occupational Health
Indian Journal of Occupational Therapy	All India Occupational Therapists Association
Indian Journal of Otolaryngology and Head and Neck Surgery	Association of Otolaryngologists of India
Indian Journal of Pediatrics	Department of Pediatrics, AIIMS
Indian Journal of Pharmacology	Indian Pharmacological Society

Cont...

Cont...	
Journal Name	**Journal Publisher**
Indian Journal of Preventive and Social Medicine	Department of P.S.M.I.M.S, Banaras Hindu University
Indian Journal of Radiology and Imaging	Indian Radiological and Imaging Association
Indian Journal of Sexually Transmitted Diseases	Indian Association For The Study of Sexually Transmitted Diseases
Indian Journal of Thoracic and Cardiovascular Surgery	Indian Association of Cardiovascular-Thoracic Surgeons
Indian Journal of Tuberculosis	Tuberculosis Association of India
Indian Pediatrics	Indian Academy of Pediatrics
J.K. Practitioner	Dr. G.M. Malik
Journal, Indian Academy of Clinical Medicine	Indian Association of Clinical Medicine
Journal of Family Welfare	Family Planning Association of India
Journal of Indian Academy of Forensic Medicine	Indian Academy of Forensic Medicine
Journal of Indian Rheumatology Association	Indian Rheumatology Association
Journal of Obstetrics and Gynecology of India	Federation of Obstetric and Gynaecological Societies of India
Journal of the Anatomical Society of India	Anatomical Society of India
Journal of Indian Association of Pediatrics Surgeons	Indian Association of Pediatrics Surgeons
Journal of Indian Society of Pedodontics and Preventive Dentistry	Indian Society of Pedodontics and Preventive Dentistry
Lung India	Indian Chest Society
Medical Journal Armed Forces India	Armed Forces Medical Services
NTI Bulletin	National Tuberculosis Institute
Trends in Biomaterials and Artificial Organs	Society For Biomaterials and Artificial Organs - India

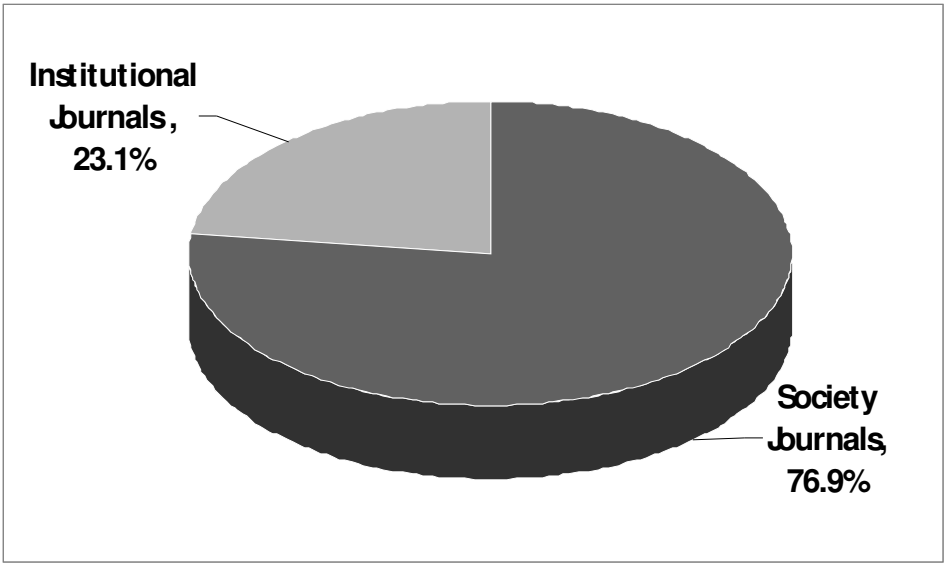


Fig. 5.17: Distribution of Two Types of Journals in medIND Project

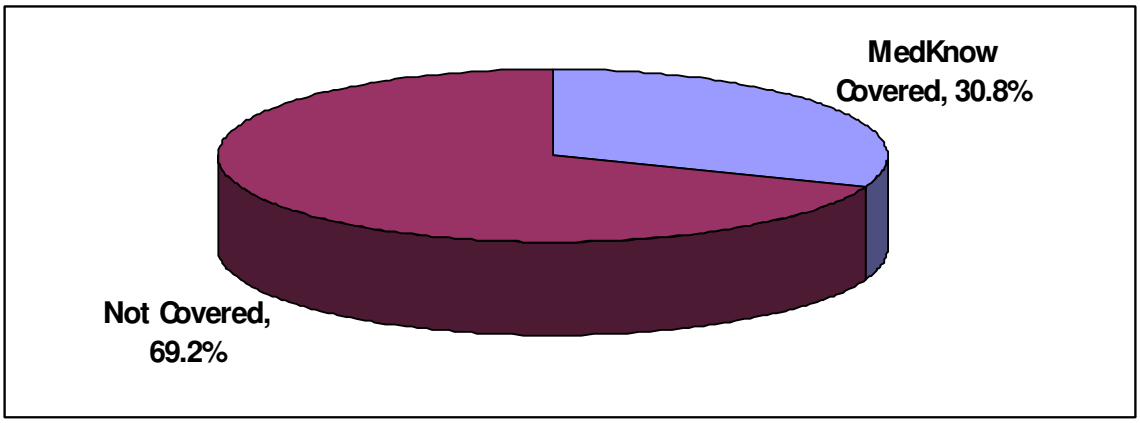


Fig. 5.18: medIND Journals covered in MedKnow Gateway

Table 5.11: Coverage of medIND Journals by Different Online Gateways

Journal Name	MedKnow Pub.	DOAJ	DOAJ Content	Google Scholar	Open J-Gate
Annals of Cardiac Anaesthesia	√	√	√	√	√
Endodontology	X	X	X	√	√
Health Administrator	X	X	X	√	√
Indian Journal of Aerospace Medicine	X	X	X	X	√
Indian Journal of Allergy Asthma and Immunology	X	X	X	X	√
Indian Journal of Anaesthesia	X	X	X	√	√
Indian Journal of Chest Diseases and Allied Sciences	X	√	X	√	√
Indian Journal of Clinical Biochemistry	X	X	X	√	√
Indian Journal of Community Medicine	√	√	√	√	√
Indian Journal of Gastroenterology	X	√	√	√	X
Indian Journal of Medical and Paediatric Oncology	X	X	X	√	√
Indian Journal of Medical Microbiology	√	√	√	√	√
Indian Journal of Medical Research	X	√	X	√	√
Indian Journal of Nephrology	√	√	√	√	√
Indian Journal of Nuclear Medicine	√	X	X	√	X
Indian Journal of Occupational and Environmental Medicine	√	√	√	√	X
Indian Journal of Occupational Therapy	X	X	X	√	√
Indian Journal of Otolaryngology and Head and Neck Surgery	X	X	X	√	√
Indian Journal of Pediatrics	X	X	X	√	X
Indian Journal of Pharmacology	√	√	√	√	√
Indian Journal of Preventive and Social Medicine	X	X	X	√	X
Indian Journal of Radiology and Imaging	√	√	√	√	√

Cont…

Cont…					
Journal Name	**MedKnow Pub.**	**DOAJ**	**DOAJ Content**	**Google Scholar**	**Open J-Gate**
Indian Journal of Sexually Transmitted Diseases	√	√	√	√	X
Indian Journal of Thoracic and Cardiovascular Surgery	X	X	X	√	√
Indian Journal of Tuberculosis	X	X	X	√	√
Indian Pediatrics	X	√	X	√	√
J.K. Practitioner	X	X	X	√	√
Journal, Indian Academy of Clinical Medicine	X	X	X	√	√
Journal of Family Welfare	X	X	X	√	√
Journal of Indian Academy of Forensic Medicine	X	X	X	√	X
Journal of Indian Rheumatology Association	X	X	X	X	√
Journal of Obstetrics and Gynecology of India	X	X	X	X	√
Journal of the Anatomical Society of India	X	X	X	X	√
Journal of Indian Association of Pediatric Surgeons	√	X	X	X	√
Journal of Indian Society of Pedodontics and Preventive Dentistry	√	√	√	√	√
Lung India	√	X	X	√	√
Medical Journal of Armed Forces India	X	X	X	√	√
NTI Bulletin	X	X	X	√	X
Trends in Biomaterials and Artificial Organs	X	√	X	√	√
Total	***12***	***14***	***10***	***33***	***31***
Coverage	***30.8%***	***35.90%***	***25.64%***	***84.62%***	***79.49%***

5.5.1 Searching and Retrieval of medIND Collection

The medIND portal provides easy navigation facility for browsing its journal collections. Here, a user can to select a journal from the list of journals available in this gateway. Then he has to specify year and issue of a journal title. Table of contents of that particular issue is then displayed on the screen. Then user can view the selected article in PDF format on screen or can download the same into his workstation. Although medIND portal does not have comprehensive search facility, metadata of medIND journal articles are maintained and searchable in the IndMED database. As subset of PubMed database, IndMED records are also searchable from online PubMed database portal. Figure 5.19 shows that online literature search either in PubMed or in IndMED database may lead to full-text access to journal literature archived in medIND portal.

Figures 5.20 to 5.22 clearly indicate that medIND collection of journals is a subset of IndMED collection, whereas IndMED collection of journals is a subset of PubMed databases. Thus, any subset of a collection can always be searchable from the parent databases. IndMED and PubMed are online bibliographic databases whereas medIND is online full-text database. Both the databases index metadata elements of journal articles archived in medIND portal thus both databases can provide full-text web-links to journal articles available in medIND gateway.

Figure 5.23 shows a screenshot of Advanced IndMED Search interface. This Figure indicates that IndMED records are searchable using different metadata elements and a combination of search is also possible using a Boolean operator. In this interface, a user can even limit his search by opting ‘restrict to medIND’ to get only full-text records available in medIND portal. Table 5.12 provides a list of metadata elements sets for the purposes of navigation, searching and displaying. For example, IndMED records can be searched by title of article, name of author, keywords, journal ISSN, source, journal title and year of publishing.

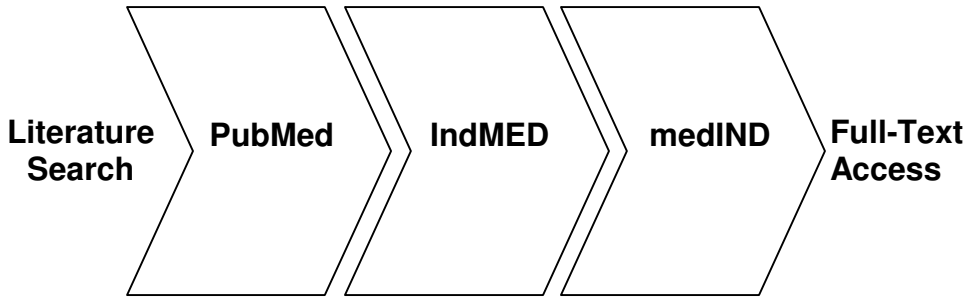


Fig. 5.19: Literature Search in PubMed and IndMED Databases to Access Full-text Articles in medIND

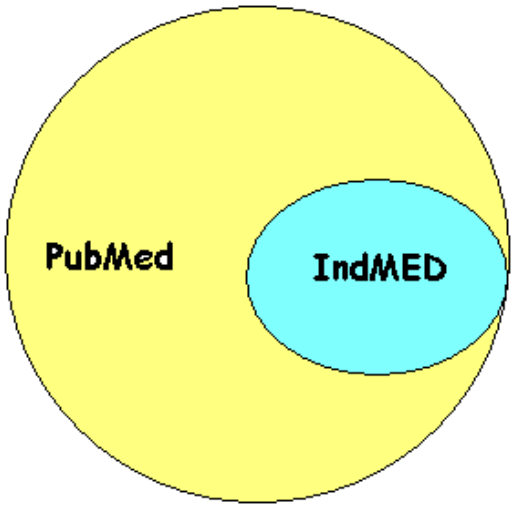


Fig. 5.20: IndMED as subset of PubMed Database

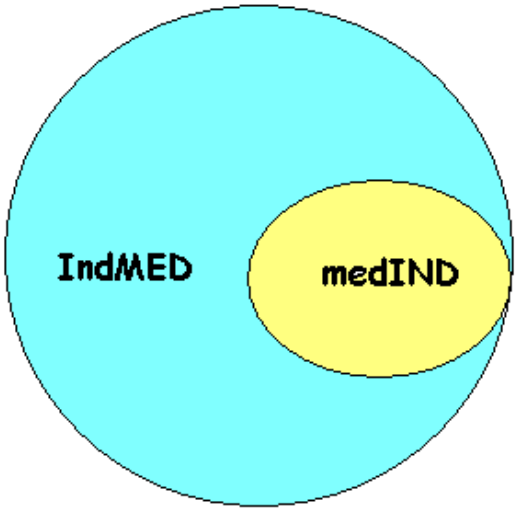


Fig. 5.21: medIND as a full-text collection of select IndMED Journals

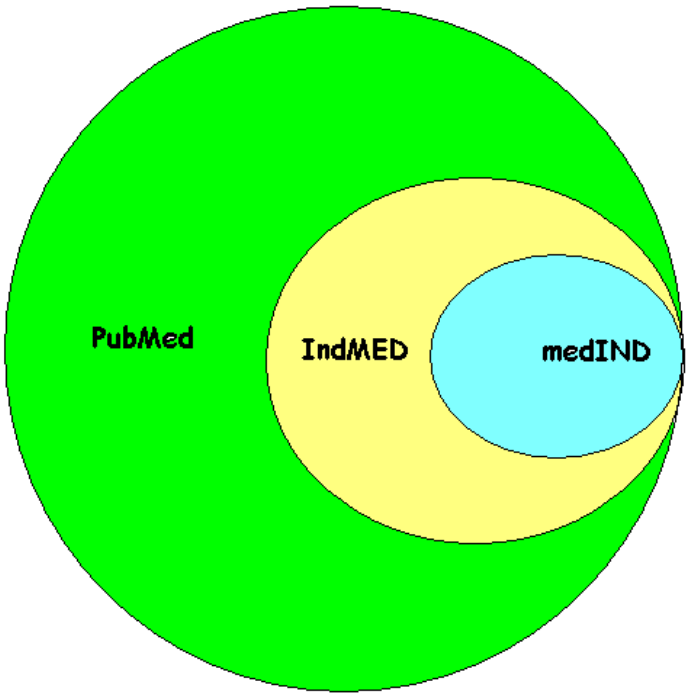


Fig. 5.22: medIND as small subset of PubMed Database

Fig. 5.23: Advanced Search Interface in IndMED, having facility to restrict within medIND articles

Table 5.12: Metadata Elements Used at medIND

Browsable Metadata	Searchable Metadata	Displayed Metadata
➢ Current Journal Issue ➢ Journal Back Issues (Name of Journal, Year, and Issue Number)	➢ Title ➢ Author ➢ Keywords ➢ ISSN ➢ Source ➢ Journal Title ➢ Year	➢ Name of Author(s) ➢ First Author's Affiliation ➢ Article Title ➢ Journal Name ➢ Year ➢ Volume No. ➢ Issue No. ➢ Pages ➢ Abstract ➢ Keywords ➢ No. of References ➢ Record Identifier ➢ Read Article (Hyperlink to Full-text) ➢ Explore Related Information (Hyperlink)

5.5.2 Sustainability Issue

When this project evolved in early 2000, it was a unique product offering to the biomedical communities in India. However, over the years medIND reaches a kind of saturation point as it could not evolve as market leader in the biomedical journal segment due to very weak marketing strategy. Nowadays e-publishing model is strategically designed to attract more international visibility and recognition. This project could not adopt a comprehensive and sustainable e-publishing model that might offer viable solutions to the problems faced by the biomedical journal publishers in India particularly a problem related to getting qualitative contents from the busy medical professionals.

5.6 Medknow Publications [www.medknow.com/journals.asp]

Medknow Publications Private Limited is a publisher of high-quality peer-reviewed scholarly open access journals in India. Medknow publishes, maintains and hosts 64+ peer-reviewed scholarly journals, mainly in the biomedical subject areas. Medknow strives for continuous improvement of visibility and accessibility of research literature emanated from the developing world. Medknow leverages latest e-publishing models in order to enhance international recognition of Indian periodicals as well as international visibility of co-publishing professional societies or institutions (Sahu, 2008; Sahu & Parmar, 2006).

Medknow provides immediate free access to the electronic editions of the journals. Publishing in Medknow journals does not cost anything to the author or author's institution. In this way Medknow promotes 'fee-less-free' open access publishing model, which is also very common to other open access initiatives in India.

Medknow Publications provide complete e-publishing solutions to its professional clients in India and abroad. Figure 5.24 provides a glimpse of publishing and online services offered by Medknow. It maintains *Journal-on-Web* – a web-based manuscript submission and peer review system that handles pre-publication and post-publication processes for journal issues, which is very innovative and new system to the Indian research journals. Other than providing e-publishing solutions, Medknow offers customized solutions, professional services and consultancy services in the areas of subscription management, advertisement management, printing and distribution.

Medknow maintains feature rich journal portal for each journal it publishes. The features of Medknow journal portals can be characterized as:

- Support to OAI-PMH compliant contents, with Dublin Core metadata and metatags for articles.
- Facility to search across the full-text of multiple journals.
- Facility to submit comments or remarks on published article.
- Ability to go to the other sites via reference and external linking.

- Linking from other sites so that visitors can reach journal site.
- Good ranking with search engines which enables people to find its journal sites.
- Support to user statistics, site structure (e.g. OpenURL), and multiple mirroring.

Medknow also offers post-print services to its clienteles that include Indexing with bibliographic databases and providing bibliographic records to indexing agencies and secondary aggregating agencies.

Figure 5.25 shows categories of Medknow clienteles availing e-publishing solutions. Its clienteles are drawn from the professional societies, academic institutions and research institutions. Most of these institutions already have scholarly periodicals in their portfolios and collaborated with Medknow to enlarge their international visibility and subscriptions. Whereas, few institutions are just starting publishing scholarly journals and collaborating with Medknow to adapt a sustainable and technologically advanced publishing model.

Figure 5.26 shows country profile of Medknow clienteles. Its clienteles are mainly drawn from India. Few institutions from countries such as Nigeria, Saudi Arabia and United Kingdom also have collaborated with Medknow for co-publishing their research journals.

Table 5.13 provides a comprehensive list of Medknow Publications hosted open access journals. In this Table a list of 64 journal titles along with their full-text availability status is shown. Co-publishing institutions having existing journals provide digitized as well as born digital contents of old volumes to Medknow for web-archiving in the respective journals' websites. Thus, very old volumes of few journals are made accessible through their websites. Medknow also helps in creation of metadata and meta-tags for digitized collection of old volumes and later these metadata records are merged with existing searchable database of published papers.

Medknow Publications specially designed its journal portals to make all Medknow journals compliant to the Open Archives Initiative-Protocol for Metadata

Harvesting (OAI-PMH). All major indexing service providers for open access journals effectively index and harvest contents of almost all Medknow journals archived in their respective portals. Table 5.14 shows a matrix of coverage of Medknow journals by online journal indexing and metadata harvesting services such as Google Scholar, Directory of Open Access Journals (DOAJ), DOAJ Content and Open J-Gate. Table 5.14 also indicates that about 85% Medknow journals are listed in DOAJ. Indexing services of DOAJ Content service, Google Scholar and Open J-Gate have outstanding coverage for Medknow journals as compared to other Indian journal gateways with 78.3%, 85% and 79.49% coverage respectively. This kind of outstanding coverage by secondary databases is due to its special initiative in providing bibliographic records to indexing agencies and secondary aggregating agencies.

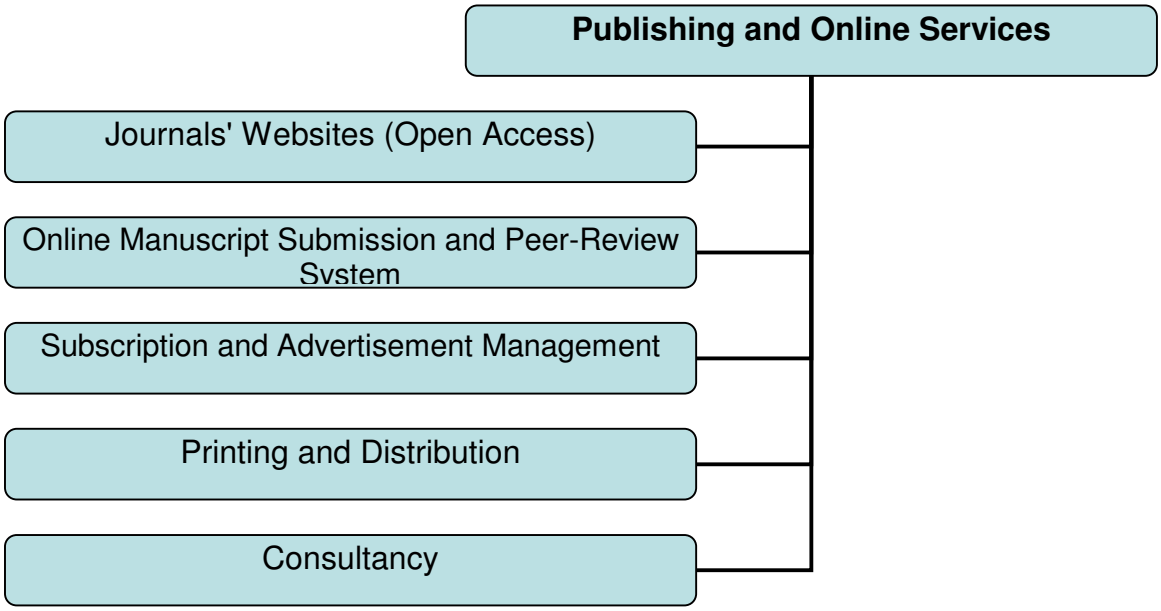


Fig. 5.24: Publishing and Online Services Provided by Medknow Publications

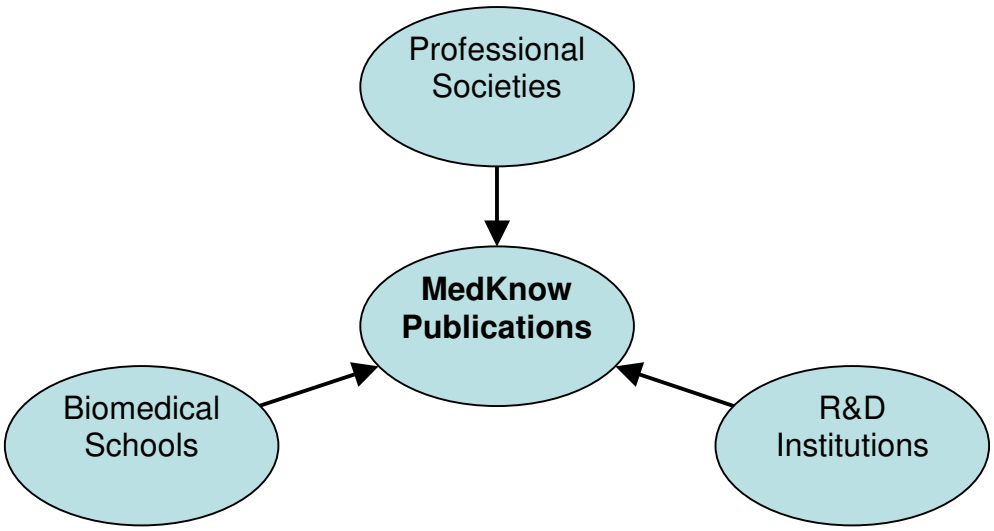


Fig. 5.25: Different Types of Clienteles for e-Publishing Solutions of MedKnow Publications

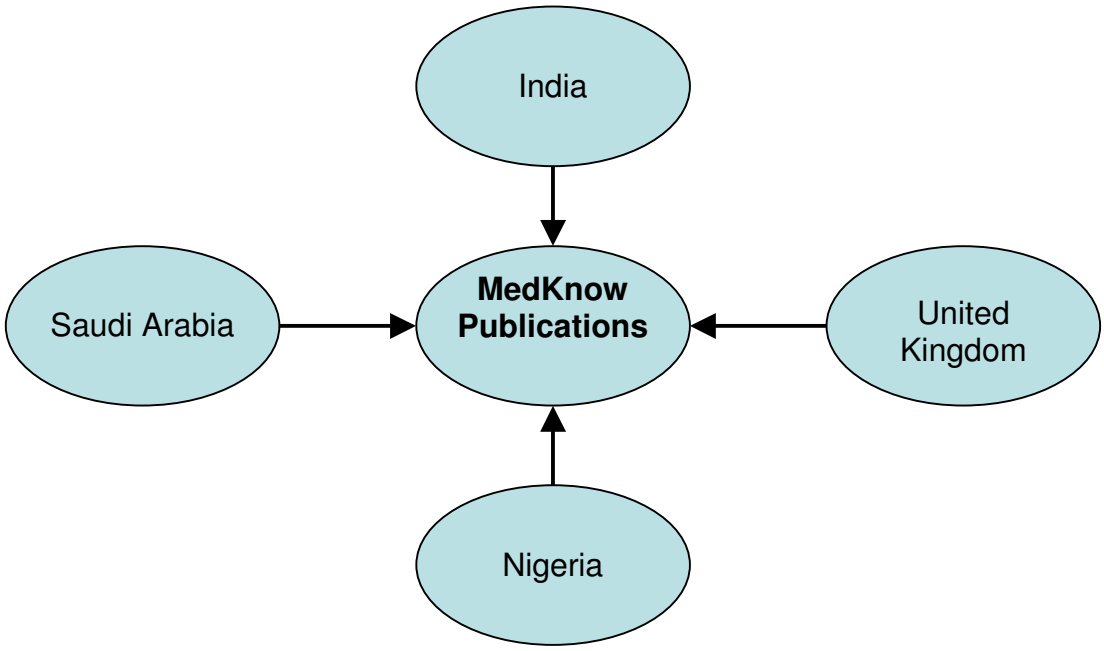


Fig. 5.26: Clienteles from Different Countries Availing e-Publishing Solutions of MedKnow Publications

Table 5.13: Medknow Publications Hosted Open Access Journals

Sr. No.	Journal Name	Full-text Availability (Year and Vol. No.)
1	African Journal of Paediatric Surgery	2007 (v.4)
2	Annals of Cardiac Anaesthesia	2005 (v.8)
3	Annals of Indian Academy of Neurology	2006 (v.9)
4	Annals of Pediatric Cardiology	2008 (v.1)
5	Annals of Thoracic Medicine	2006 (v.1)
6	Annals of Tropical Medicine and Public Health	2008 (v.1)
7	Asian Journal of Pharmaceutics	2008 (v.2)
8	Asian Journal of Transfusion Science	2007 (v.1)
9	CytoJournal	2008 (v.5)
10	Hepatitis B Annual	2004 (v.1)
11	Indian Journal of Cancer	2003 (v.39)
12	Indian Journal of Community Medicine	2007 (v.32)
13	Indian Journal of Critical Care Medicine	2003 (v.7)
14	Indian Journal of Dental Research	2006 (v.17)
15	Indian Journal of Dermatology	2005 (v.50)
16	Indian Journal of Dermatology, Venereology, and Leprology	2001 (v.56)
17	Indian Journal of Human Genetics	2002 (v.8)
18	Indian Journal of Medical Microbiology	2001(v.19)
19	Indian Journal of Medical Sciences	2001 (v.55)
20	Indian Journal of Nephrology	2007 (v. 17)
21	Indian Journal of Nuclear Medicine	2008 (v.23)
22	Indian Journal of Occupational and Environmental Medicine	2003 (v.7)
23	Indian Journal of Ophthalmology	1975 (v.23)
24	Indian Journal of Orthopaedics	2007 (v.41)
25	Indian Journal of Palliative Care	2004 (v.9)

26	Indian Journal of Pathology and Microbiology	2008 (v.51)
27	Indian Journal of Pharmaceutical Sciences	2006 (v.68)
28	Indian Journal of Pharmacology	1969 (v.1)
29	Indian Journal of Plastic Surgery	2001 (v.34)
30	Indian Journal of Psychiatry	2005 (v.47)
31	Indian Journal of Psychological Medicine	2008 (v.30)
32	Indian Journal of Radiology and Imaging	1999 (v.9)
33	Indian Journal of Sexually Transmitted Diseases and AIDS	2007 (v. 28)
34	Indian Journal of Surgery	2003 (v.65)
35	Indian Journal of Urology	2005 (v.21)
36	International Journal of Diabetes in Developing Countries	2001 (v.21)
37	International Journal of Green Pharmacy	2008 (v.2)
38	International Journal of Shoulder Surgery	2007 (v.1)
39	International Journal of Yoga	2007 (v.1)
40	Journal of Cancer Research and Therapeutics	2005 (v.1)
41	Journal of Carcinogenesis	2008 (v.7)
42	Journal of Conservative Dentistry	2007 (v.10)
43	Journal of Cutaneous and Aesthetic Surgery	2008 (v.1)
44	Journal of Cytology	2008 (v.25)
45	Journal of Emergencies, Trauma and Shock	2008 (v.1)
46	Journal of Gynecological Endoscopy and Surgery	2009 (v.1)
47	Journal of Human Reproductive Sciences	2008 (v.1)
48	Journal of Indian Association of Pediatric Surgeons	2005 (v.10)
49	Journal of Indian Society of Pedodontics and Preventive Dentistry	2003 (v.21)
50	Journal of Laboratory Physicians	2009 (v.1)

51	Journal of Medical Physics	2006 (v.31)
52	Journal of Minimal Access Surgery	2005 (v.1)
53	Journal of Oral and Maxillofacial Pathology	2007 (v.11)
54	Journal of Pediatric Neurosciences	2006 (v.1)
55	Journal of Postgraduate Medicine	1980 (v.26)
56	Journal of Sexual Medicine, Andrology and Gender	2009 (v.1)
57	Lung India	2008 (v.25)
58	Medical Law Cases for Doctors	2008 (v.1)
59	Mens Sana Monographs	2003 (v.1)
60	Neurology India	1999 (v.47)
61	Noise and Health	1998 (v.1)
62	Saudi Journal of Gastroenterology	2006 (v.12)
63	The Journal of Indian Prosthodontic Society	2005 (v.5)
64	Urology Annals	2009 (v.1)

Table 5.14: Coverage of Medknow Journals by Different Online Gateways

Journal Name	DOAJ	DOAJ Content	Google Scholar
African Journal of Paediatric Surgery	√	√	√
Annals of Cardiac Anaesthesia	√	√	√
Annals of Indian Academy of Neurology	√	√	√
Annals of Pediatric Cardiology	√	√	√
Annals of Thoracic Medicine	√	√	√
Annals of Tropical Medicine and Public Health	*X*	*X*	*X*
Asian Journal of Pharmaceutics	√	√	*X*
Asian Journal of Transfusion Science	√	√	√
CytoJournal	√	√	√
Hepatitis B Annual	√	√	√
Indian Journal of Cancer	√	√	√
Indian Journal of Community Medicine	√	√	√
Indian Journal of Critical Care Medicine	√	√	√
Indian Journal of Dental Research	√	√	√
Indian Journal of Dermatology	√	√	√
Indian Journal of Dermatology, Venereology and Leprology	√	√	√

Indian Journal of Human Genetics	√	√	√
Indian Journal of Medical Microbiology	√	√	√
Indian Journal of Medical Sciences	√	√	√
Indian Journal of Nephrology	√	√	√
Indian Journal of Nuclear Medicine	X	X	√
Indian Journal of Occupational and Environmental Medicine	√	√	√
Indian Journal of Ophthalmology	√	√	√
Indian Journal of Orthopaedics	√	√	√
Indian Journal of Palliative Care	√	√	√
Indian Journal of Pathology and Microbiology	√	√	√
Indian Journal of Pharmaceutical Sciences	√	√	√
Indian Journal of Pharmacology	√	√	√
Indian Journal of Plastic Surgery	√	√	√
Indian Journal of Psychiatry	√	√	√
Indian Journal of Psychological Medicine	X	X	X
Indian Journal of Radiology and Imaging	√	X	√
Indian Journal of Sexually Transmitted Diseases and AIDS	√	√	X
Indian Journal of Surgery	X	X	√
Indian Journal of Urology	√	√	√
International Journal of Diabetes in Developing Countries	√	√	√
International Journal of Green Pharmacy	√	X	√
International Journal of Shoulder Surgery	√	√	√
International Journal of Yoga	√	√	X
Journal of Cancer Research and Therapeutics	√	√	√
Journal of Carcinogenesis	√	√	√
Journal of Conservative Dentistry	X	X	X
Journal of Cutaneous and Aesthetic Surgery	X	X	X
Journal of Cytology	√	√	√
Journal of Emergencies, Trauma and Shock	√	X	X
Journal of Human Reproductive Sciences	√	√	√
Journal of Indian Association of Pediatric Surgeons	√	√	√
Journal of Indian Society of Pedodontics and Preventive Dentistry	X	X	√
Journal of Medical Physics	√	√	√
Journal of Minimal Access Surgery	√	√	√
Journal of Oral and Maxillofacial Pathology	√	√	√
Journal of Pediatric Neurosciences	√	√	√
Journal of Postgraduate Medicine	√	√	√
Lung India	X	X	√

Medical Law Cases for Doctors	X	X	X
Mens Sana Monographs	√	√	√
Neurology India	√	√	√
Noise and Health	X	X	√
Saudi Journal of Gastroenterology	√	√	√
The Journal of Indian Prosthodontic Society	√	√	√
Total	**51**	**47**	**51**
Coverage	**85%**	**78.3%**	**85%**

5.6.1 Searching and Retrieval of Medknow Collection

Medknow portal provides easy navigation facility for browsing its journal collections. Medknow maintains separate journal portal for each journal. Here, a user can select a journal from the list of journals available in this gateway. Then that particular journal portal is opened up in a new window. There he has to navigate the journal issues by selecting a specific year and then specific issue number. Table of contents of that particular issue is then displayed on the screen. Then user can view either abstract or full-text content of selected article in HTML and PDF formats on screen or can download the same into his workstation. Medknow portal management system generates HTML file from the XML file it maintains for a full-text article.

Medknow portal also provides facility to search across the full-text of multiple journals. For this purpose, searchable metadata are only author and keyword. After a successful search query, portal system generates search result displaying metadata elements such as article title, name of author(s), journal name, year, volume number, issue number, article type, abstract and hyperlinks to full-text content in HTML and PDF. Medknow collections are also searchable from different secondary databases such as Google Scholar, Open J-Gate and DOAJ Content. Table 5.15 provides a list of metadata elements sets used in Medknow portal for the purposes of navigation, searching and displaying. For example, Medknow bibliographic records can be searched by name of author and keywords, whereas Medknow journals can be searched by subject, ISSN, etc.

Table 5.15: Metadata Elements Used at Medknow

Browsable Metadata	Searchable Metadata	Displayed Metadata
➢ Table of Contents of Current Journal Issues ➢ Journal Back Issues (Name of Journal, Year, and Issue Number)	➢ Author ➢ Keywords ➢ ISSN ➢ Journal Name	➢ Article Title ➢ Name of Author(s) ➢ Journal Name ➢ Year ➢ Volume No. ➢ Issue No. ➢ Abstract ➢ Article Type ➢ Read Article (Hyperlink to Full-text HTML and PDF)

Medknow initiative is characterized by its future-readiness which is an ability to adapt with the changing norms of e-publishing and changing e-publishing technologies. Medknow is also very pro-active to achieve good ranking with search engines that enables people to find its journal sites. Before any crawler from a journal indexing service visit Medknow journals, Medknow voluntarily provides metadata information to those metadata harvesters and journal indexing services on its recent additions. This way it Medknow is very successful open access initiative in India having an impressive international visibility of its journals.

5.7 Udbodhan Magazine [www.udbodhan.org/article.htm]

The Udbodhan is more than a century old Bengali magazine, published by Ramakrishna Math Udbodhan in Kolkata. Inspired by Swami Vivekananda, Swami Trigunatitananda started this monthly magazine in January 1899. This magazine covers a wide range of subjects such as spirituality, comparative religious thoughts, social work, social development, youth development, Indian literature, Indian culture and heritage.

The publisher took up a huge task of digitizing this esteemed magazine to make available this important collection of Ramakrishna-Vivekananda literature to the Bengali knowing computer users all over the world, which took a few years to complete this digitization project, as indicated in Figure 5.27. In September 2005, Ramakrishna Math released the e-version of all the *Udbodhan* articles in 21 CD-ROMs covering contents published in its first centenary. A portion of the digitized contents is web-enabled in the magazine's web portal. Udbodhan magazine portal provides open access to full-text contents of this magazine since July 2000 issue. Inaugural issue of this magazine published in January 1899 is also made online in this portal. Presently, current issues of this magazines are born-digital and web-enabled immediately publishing its print counterpart. Although this magazine is freely available online, its print version is subscription based. So is the CD-ROM collection of digitized articles.

Ramakrishna Math is also engaged in digitizing other important literature collections such as *Prabuddha Bharata*, published by Advaita Ashrama – a wing of Ramakrishna Math, and other important works on Ramakrishna-Vivekananda ideology and Vedanta.

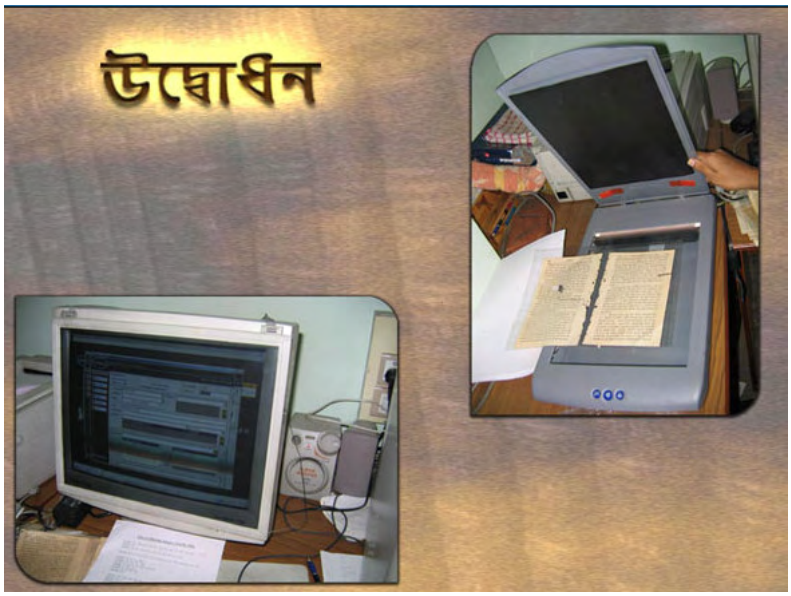


Fig. 5.27: Digitization Work in Progress for Udbodhan Collection

5.7.1 Searching and Retrieval in Udbodhan Centenary CD-ROM Collection

Udbodhan Centenary CD-ROM collection provides easy navigation facility for browsing its magazine collection. Using *Image Viewer*, any offline user can browse the articles in the Udbodhan Patrika 100 Years collection. In this mode the user first selects the year of publication, then issue of the publication. As the Patrika is fortnightly serial, in each Bengali calendar month two issues are published. When user selects a particular issue of the magazine, table of contents of that issue appears on the screen. Then user selects an article of his choice. The digitized text of the article appears in another window, where user can read the article page-wise. Thus browsing an article is five clicks away from the user - (i) select Image Viewer, (ii) select year range (e.g. 1380-1389), (iii) select particular year (e.g. 1385), (iv) select particular issue (e.g. Baishakh-1), and (v) select particular title of the article. In the image viewer mode, as shown in Figure 5.28, browsable metadata elements are displayed in Bengali language.

Table 5.16 provides a list of metadata elements sets used in Udbodhan Centenary collection for the purposes of navigation, searching and displaying. For example, this CD-ROM collection has simple and advanced search options. A search can be carried out using one or more searchable metadata elements such as category, subject, article title, author type and author name. Figure 5.29 shows the simple search interface in Udbodhan Collection. The search result is displayed in transliterated English that include CD label (or CD number), title of article, volume number, issue and year of publishing.

5.7.2 Importance of an Offline Digitized Collection

Figure 5.1 in the Section 5.1 shows that CD-ROM collection is a viable option for dissemination of offline digitized contents. Offline collections are very useful to the institutions having weak or no internet connectivity, particularly to the community libraries and school libraries. The restrictions in intellectual property rights also hamper online dissemination of digitized collections, unless the publishers have efficient digital rights management mechanism. The Udbodhan

Centenary CD-ROM collection is included in this study in recognition to the importance of this type of offline digitized contents in the processes of teaching, learning and self-actualization.

Table 5.16: Metadata Elements Used in Medknow

Browsable Metadata	Searchable Metadata	Displayed Metadata
➢ Range of Years ➢ Year ➢ Month ➢ Issue Number ➢ Table of Contents of the Selected Issue ➢ (Displayed in Bengali)	➢ Category ➢ Subject ➢ Title ➢ Author Type ➢ Author Name	➢ CD Label (or CD number) ➢ Title of Article ➢ Volume Number ➢ Issue ➢ Year of Publishing ➢ (Displayed in Transliterated English)

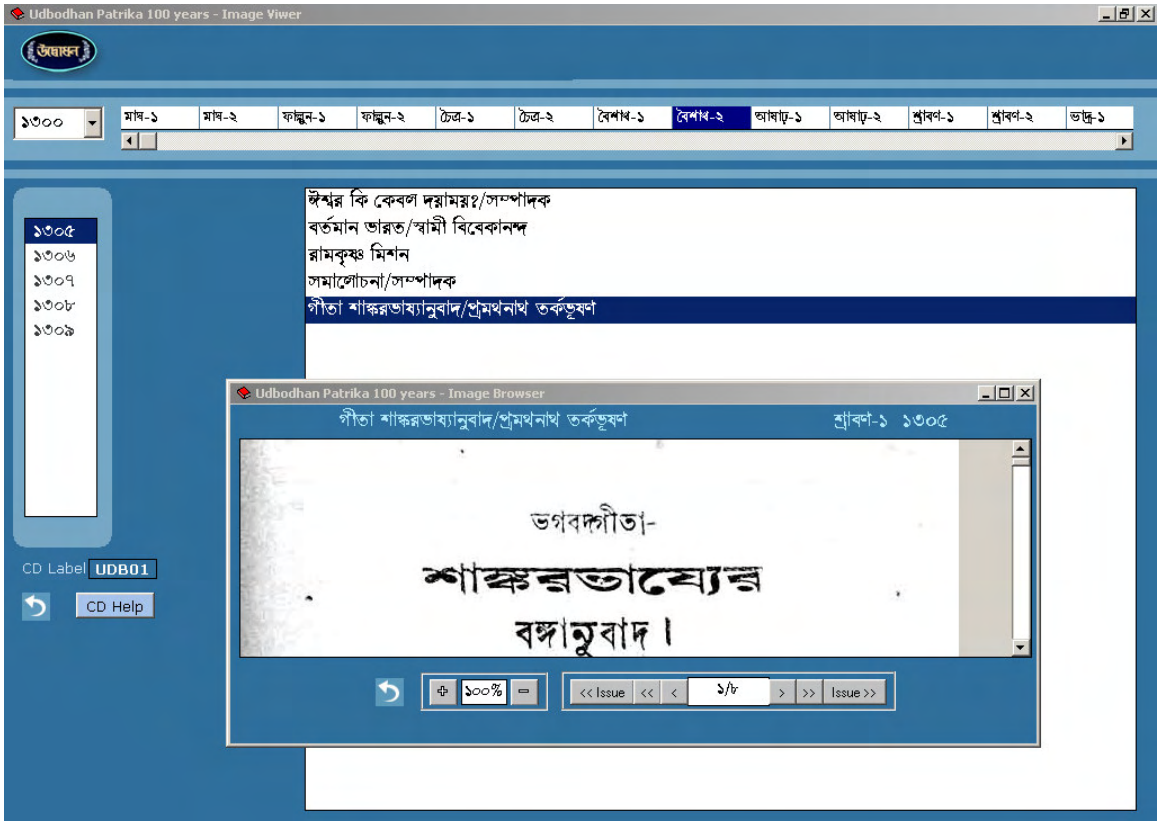


Fig. 5.28: Image Viewer – A Navigation Tool for Udbodhan Centenary CD-ROM Collection

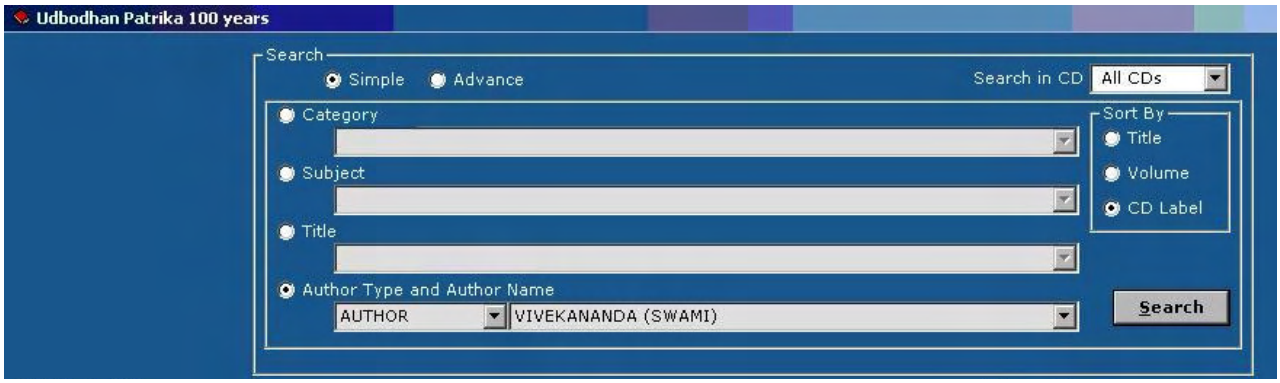


Fig. 5.29: Search Engine for Udbodhan Centenary CD-ROM Collection

5.8 Some Observations

India is spearheading open access movement in Asia with participation of the national level institutions such as Indian National Science Academy, Indian Academy of Sciences and National Informatics Centre as well dynamic private players such as Medknow Publications. In addition to journal gateways, many individual academic journals are also now available through open access channel to increase the international visibility of Indian research literature. In this chapter, different journal content digitization initiatives are evaluated in terms of their availability in the open access indexing and harvesting services. Unfortunately, some of the digitization initiatives did not have a roadmap for implementing OAI-PMH-enabled web portals. However, early adoption of OAI-PMH-based system is required to make the Indian scholarly periodicals indexed and harvested in majority of indexing services. Moreover, availability of Indian journals in e-publishing channels needs to be accelerated in order to make them more accountable to its stakeholders such as members of professional societies publishing the journals as well as achieving international readership.

5.9 References

Canessa, E. & Zennaro, M., eds. (2008). *Science dissemination using open access: a compendium of selected literature on open access*. Trieste: Abdus Salam International Centre for Theoretical Physics.

European Commission (2008). *Open access: opportunities and challenges*. Brussels: Directorate-General for Research, European Commission.

Ghosh, S.B. & Das, A.K. (2007). Open access and institutional repositories - a developing country perspective: a case study of India. *IFLA Journal*, 33(3), 229-250.

Mujoo-Munshi, U. (2003). Building digital resources: creating facilities at INSA. *The International Information & Library Review*, 35(2-4), 281-309.

Sahu, D.K. (2008). Gaining impact, readers and authors through fee-less-free dissemination: an experiment with open access. In: *Brainstorming Meet on Open Access, FLOSS and Copyright Law for Scholarly Communication and Literary Work*, 26 April 2008, New Delhi, India.

Sahu, D.K. & Parmar, R. (2006). The position around the world: open access in India. In: *Open access: key strategic, technical and economic aspects*. Oxford, UK: Chandos Publishing (Oxford) Limited.

Sen, B. K. (2002). Growth of scientific periodicals in India (1788-1900). *Indian Journal of History of Science,* 37(1), s1-s46.

Singh, S., Gaba, S.K. & Pandita, N. (2004). Architecture and building of Medical Digital Library at NIC [of India]: What exists and what is required for MeDLib@NIC?. In: *ICDL 2004 proceedings*. New Delhi: TERI.

CHAPTER 6

Conclusion and Recommendations

In today's environment, hoarding knowledge ultimately erodes your power. If you know something very important, the way to get power is by actually sharing it.

Joseph Badaracco

Professor of Business Ethics at Harvard Business School

6.1 Prologue

Digitization of library materials has quickly become common practice in a great many institutions. It provides an excellent opportunity to widely disseminate our documentary heritages and greatly increases access to library collections of rare documents as well as current research literature. In this connection I would like to provide an anecdote based on a true story:

> *One professor of Jawaharlal Nehru University (JNU) went to Kolkata from New Delhi to consult National Library of India (NLI) and locate a specific book there. NLI neither has any web-based online public access catalogue (OPAC) nor an email-based service to inform a distant user about availability of a document. Fortunately, he came to know about availability of the book in NLI collections from his Kolkata-based friend. After two days of follow up, he could locate the book in the library and took photocopy of about 50% of the said book, although he needed full copy of the book for his research purpose. This was an out of copyrighted book, but lack of understanding by NLI staff on copyright norms forced him to settle for a partial content. After returning back to JNU, he asked his colleague for obtaining remaining part of the book from the NLI. Instead of going to NLI, his colleague referred him to Digital Library of India (DLI) portals. Both of*

them found full text content of the book at a DLI site within a few minutes and took print outs on the same day.

Digital Library of India has digitized many historically important documents and those are now available in DLI portals. Many researchers and scholars need to consult old literature for their advanced studies. However, many researchers in India either are not competent in Internet searching or not aware of online resources originated in India. Thus, when we see usage statistics of any digital library or digital archive site originated from India including DLI sites, we observe that the top ranking country is either United States or any other developed country in terms of number of visitors in a particular period from a country, but not India. This reflects the fact that Indian researchers and scholars need training on functional information literacy and digital literacy, which is very crucial at this digital age.

6.2 Summary of Findings

Indian digitization initiatives as discussed in the Chapters 2 to 5 aim at producing a vast amount of digitized documents pertaining to different forms of recorded human knowledge, ranging from the rare manuscripts to current research literature. Digital Library of India is the largest digitization initiative in India spreading across states of India and involving over ninety organizations to ensure several thousands of rare books written in Indian languages as well as non-Indian languages are accessible through Internet channel. Chapter 2 critically appraises different aspects of the DLI project such as collaboration pattern, digitization processes, metadata practices, information retrieval, etc.

The higher education system in a country produces advanced research literature in the form of theses and dissertations in both thrust areas of the country as well as in the emerging academic fields. The researchers in the country need to consult results of past researches in order to expand frontiers of knowledge. Theses and dissertations are constant in demand for our academic research, although availability of the same is the matter of concern. Chapter 3 critically

appraises different aspects of digitization work for theses and dissertations in higher learning institutions in India, more particularly initiatives such as Vidyanidhi Digital Library project, ETD@IISc and OpenMed@NIC.

Documentary heritage collections in the memory institutions in the country are on the verge of extinction. Chapter 4 critically appraises digitization programmes of two national level institutions covering the documentary heritage collections across states of India. While Indira Gandhi National Centre for the Arts (IGNCA) is engaged in multimedia documentation and digitization of multilingual documents in the areas of Indian culture, art and heritage, National Mission for Manuscripts focuses on most valuable collections of rare manuscripts covering many subject areas.

Journal literature is an important kind of research communication, where scholars publish their research findings. Indian journals can be found in almost every subject field – making a value addition to the world research literature. However, Indian journals were not available to the international research communities due to the weaker distribution channel. Indian journals as well as research papers published by Indian scholars in these journals were less visible to the international research communities. However, the research communication process has been drastically changed with the adoption of modern ICT-enabled publishing environment by the Indian journal publishers. Most of the publishers of Indian research journals belong to non-profit making scientific societies or institutions. Thus, open access to Indian journal literature was one of the viable options for making Indian journals internationally visible and available. Chapter 5 critically appraises digitization initiatives of some scientific institutions in India to make the retrospective volumes of Indian journals available. This chapter covers four major Indian open access journal initiatives, namely, e-Journals@INSA, Indian Academy of Sciences published journals, medIND@NIC, Medknow Publications.

6.2.1 In Search of Theoretical Frameworks on Public Access to Digitized Collections

Indian digitization projects as well as open access initiatives have experimented with different phases of transformation. For any kind of material, these transformation phases are almost common with certain degrees of variations or deviations. In this section, a range of theoretical models is proposed in order to set them into a theoretical framework for best describing a digitization initiative in India.

a) Model 1: Public Access to Digitized Collections

The Model 1 depicts four broad phases of transformation. In the first phase of this model, the document collections have undergone through the process of digitization. In this phase, neither a systematic storage nor a retrieval system is available. Thus, the digitized collections are not available either in intranet or Internet channel.

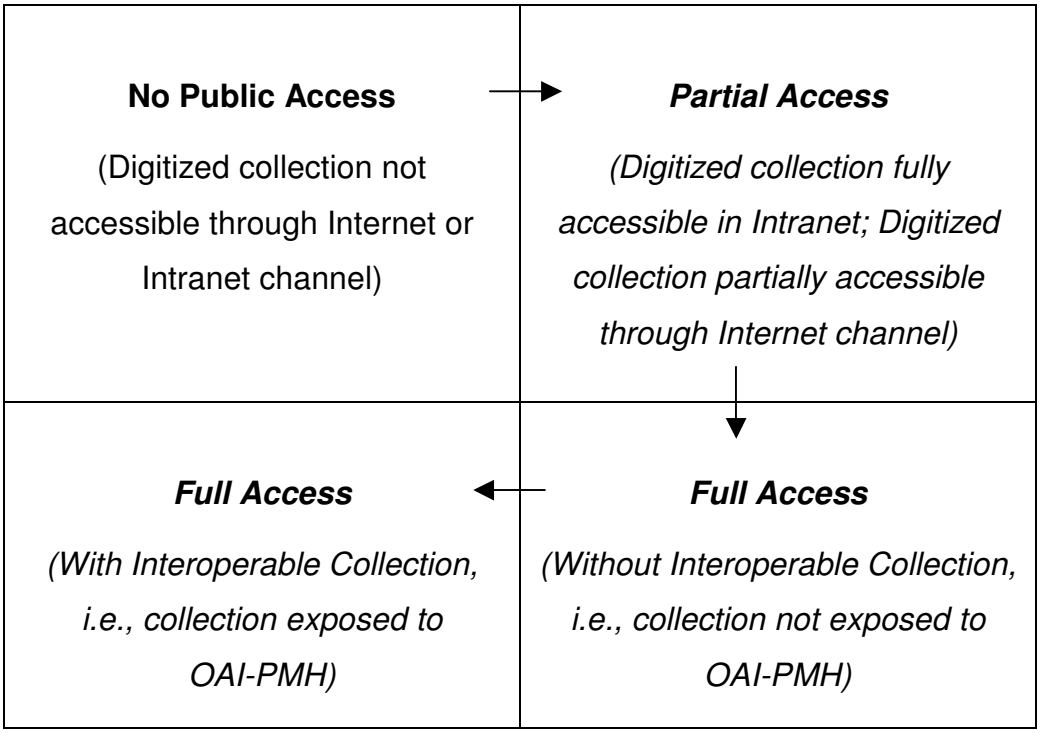
In the second phase of Model 1, the whole collection is digitized and information storage and retrieval system is functioning. The whole digitized collection is made accessible in intranet. In addition to that a certain portion of whole digitized collection is made publicly accessible through Internet channel. The partially visible collection is maintained due to many problems as perceived by the implementing agency such as bandwidth limitation, copyright restrictions, lack of high capacity Webservers, and delays in quality checking before web-enablement. Best known example for this phase is the Kalasampada: DL-RICH project of IGNCA. In this project, less than 5% of whole collections are available in public domain and remaining portion (over 95%) is inaccessible to the Internet-based users.

In the third phase of Model 1, the whole digitized collection of the digital library is made accessible through both intranet and Internet channel. The whole collection can be browsed or navigated through the digital library Webportals of that initiative. However, the digitized collection is not interoperable, i.e., collection is not exposed to OAI-PMH (Open Archives Initiative-Protocol for Metadata

Harvesting), thus, digitized collection is not indexed by any secondary indexing service. Best known example for this phase is the Digital Library of India project. The DLI portals do not support OAI-PMH-based metadata indexers that could make great accessibility of the portals.

In the fourth and final phase of Model 1, the whole digitized collection of the digital library is made accessible through both intranet and Internet channel. The whole collection can be browsed or navigated through the digital library Webportals of that initiative. The digitized collection is also interoperable, i.e., collection is exposed to OAI-PMH, and thus, the digitized collection is indexed by many secondary indexing services. Best known example for this phase is the Vidyanidhi Digital Library project. Vidyanidhi portal, having full-text ETD collection from Indian universities, has been developed using DSpace open source software that supports interoperability. OAI-PMH-based metadata indexers obtain metadata information from this portal and enhance its accessibility through online indexing services from different portals, such as OpenDOAR, Google Scholar, CASSIR, etc.

Model 1: Public Access to Digitized Collections



b) Model 2: Growth of Open Access Journals in India

The Model 2 depicts four broad phases of transformation in the digitization work for Indian journal collections. In the all four phases in this model, digitization project has been completed and digitized articles are made web-enabled. The information storage and retrieval system is also implemented there.

The first phase of this model is marked as Static Growth phase as digitized article collection has become static and born digital articles are added very rarely. Best known example for this phase is e-Journals@INSA, implemented by Indian National Science Academy. In this portal digitized journal articles of old volumes are only archived and current volumes of post digitization period are almost unavailable there.

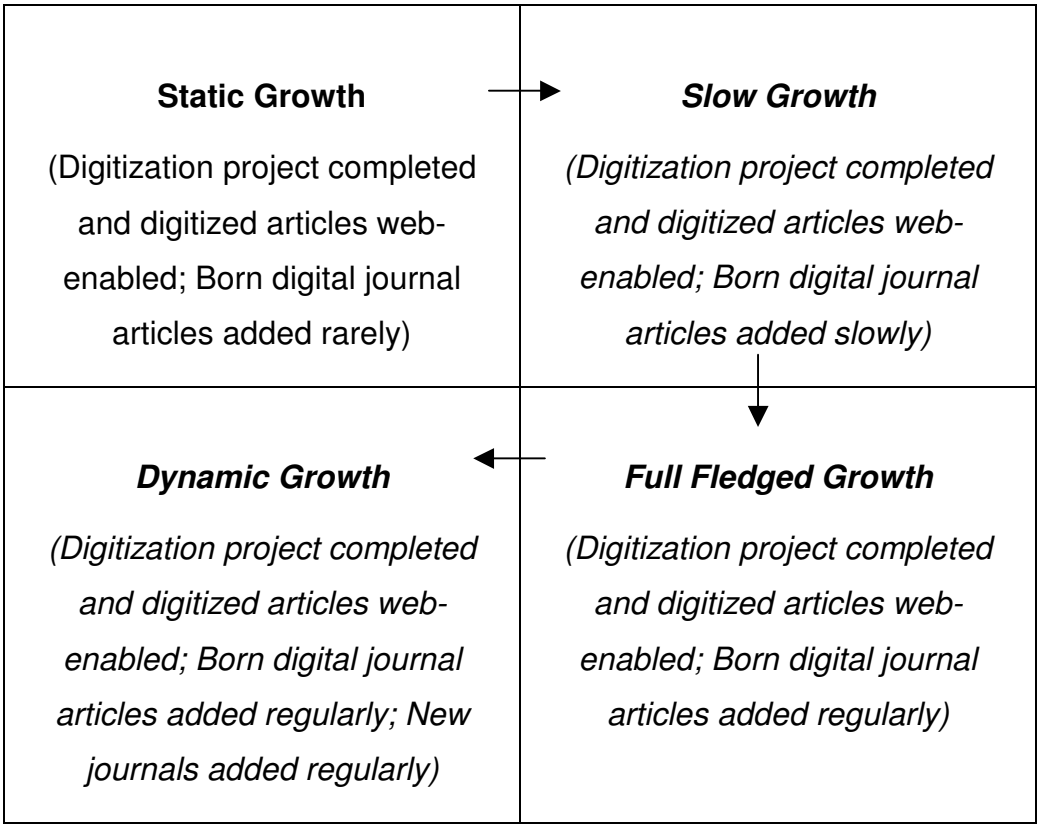
The second phase of this model is marked as Slow Growth phase as digitized as well as born digital articles are slowly added in a portal. Best known example for this phase is medIND@NIC, implemented by Bibliographic Services Division of National Informatics Centre. In this portal digitized journal articles of old volumes of Indian biomedical periodicals are archived and current volumes of post digitization period are slowly added there.

The third phase of this model is marked as Full Fledged Growth phase. In this phase digitized article collection is almost fully archived, and born digital articles are very regularly added in the journals gateway. Best known example for this phase is open access journals published by Indian Academy of Sciences (IAS). In this gateway digitized journal articles of old volumes of IAS journals are almost fully archived and current volumes of post digitization period are regularly added there. There is no time gap between print journal publication and web-enablement of current issue of an IAS journal. In fact most of the journal issues are published online earlier than their print counterparts. However, number of journal titles published by the Academy remains same.

The fourth and final phase of this model is marked as Dynamic Growth phase. In this phase, the journal gateway provides full access to digitized articles as well as born digital articles of current volumes of journals. A number of new journal titles

are also added in this gateway at regular interval through the process of new journal acquisition, bifurcation or initiation. Best known example for this phase is the Medknow publications. Medknow is new age digital publisher, publishing open access journals in collaboration with biomedical professional societies and institutions. Most of the journal issues of Medknow journals are published online earlier than their print counterparts.

Model 2: Growth of Open Access Journals in India



c) Model 3: Open Access to e-Theses Collections in India

The Model 3 depicts four broad phases of transformation of the digitization work for thesis and dissertation collections. In the all four phases of this model, theses and dissertations are either digitized or available as born digital objects.

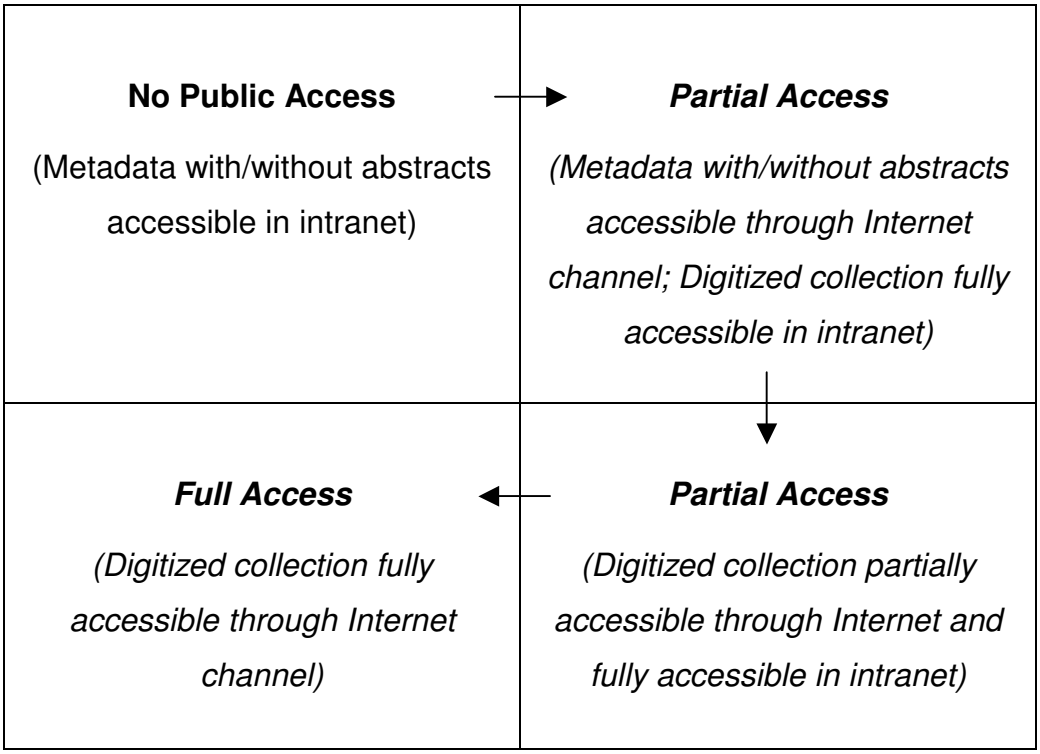
In the first phase of this model, only metadata information with or without abstracts of thesis collection is made available in intranet. In the second phase, only metadata information with or without abstracts of thesis collection is made

available through Internet channel whereas full text ETD collection is made available in intranet for the internal users. Best examples in this phase are national initiatives such as CSIR e-Thesis and institutional initiatives of IIT Chennai, IIT Mumbai, IIT Kanpur, etc.

In the third phase of this model, a partial full text collection is made accessible through Internet channel, whereas same is fully accessible in intranet for the internal users. The best example of this phase is ETD collection of IIT Delhi which is partially accessible through its institutional repository - ePrints@IIT Delhi and fully accessible in campus-wide Intranet. In the fourth and final phase of this model, entire ETD collection is made accessible in Internet through a web-based ETD repository. The best examples of this kind of collections are Vidyanidhi ETD repository and ETD@IISc repository.

Figure 6.1 depicts a chain of transitional phases from no online public access to full-text open access to ETD collection, based on Model 3.

Model 3: Open Access to e-Theses Collections in India



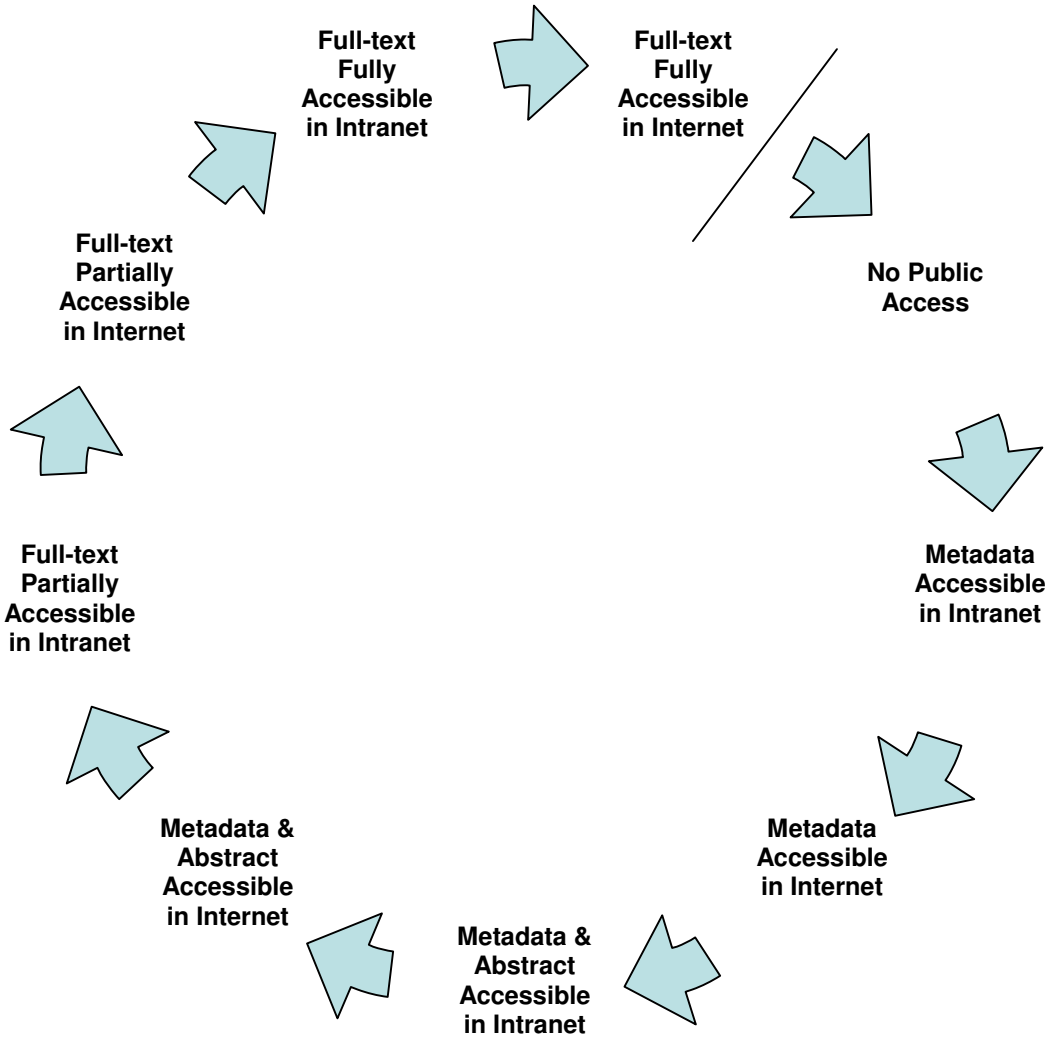


Fig. 6.1: Transition from No Online Public Access to Full-text Open Access to ETD Collection

d) Model 4: Sustainability of Digitization Projects

The Model 4 takes care of sustainability as well as resource mobilization factors in Indian digitization projects. A digitization initiative in India is characterized primarily by two types of approaches, one is bottom-up approach and another

one is top-down approach. In the bottom up approach, an Indian institution realizes the necessity of a digitization and/or open access project to make accessible its published literature through digital media, thus, the institution undertakes a digitization or open access project mobilizing its own resources or obtaining financial grants from an external agency. A number of institutional repositories have been implemented in Indian institutions, where bottom up approach is mostly followed. The first phase in the Model 4 is mostly based on bottom up approach having institutionally-supported or self-sponsored digitization projects.

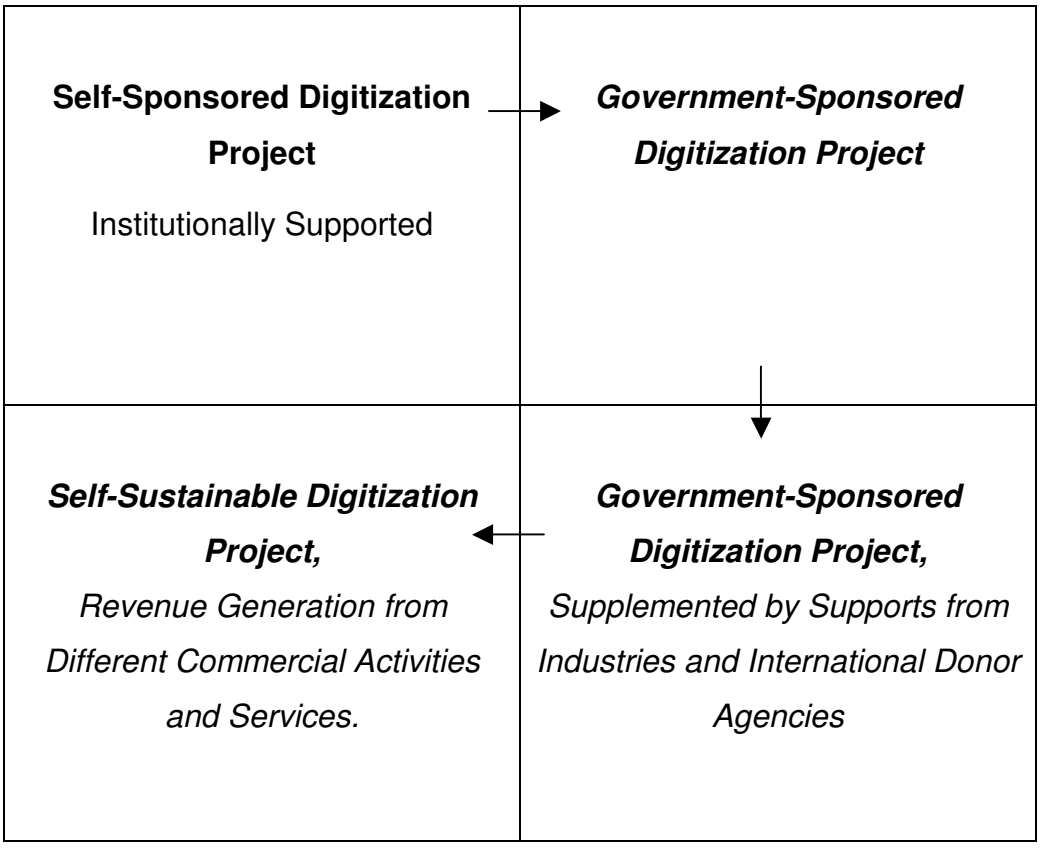
In the top down approach, an apex agency in the country envisages a national digitization project in partnership with an array of institutions of a specific type. Phases two and three projects in this Model mostly follow top-down approach, where a central agency is responsible for project planning, national coordination and resource mobilization. In the phase two of this Model, the digitization project is fully funded by a government department. Best examples for this phase are: MCIT supported digitization projects implemented by IGNCA, and Ministry of Culture supported manuscript digitization project implemented by National Mission for Manuscripts. The main problem in the project of this phase is that the project may not sustain for a long run after the completion of project period, unless the implementing agency is very keen to continue the digital library services.

In the third phase of this model, the project continues to receive government grants, and additionally it generates resources from industries and national/ international donor agencies. As this phase generates more resources and ends up with multi-stake partnership, sustainability of the digitization project or digital library services is widened in a certain extent. The best examples of this phase of initiatives are Vidyanidhi Digital Library project and Digital Library of India project. In Vidyanidhi Digital Library project, Microsoft India and Ford Foundation have been contributed in addition to support from the NISSAT, DSIR. In DLI project, United States National Science Foundation has also been contributed in addition to support from the MCIT.

The fourth and final phase of this model takes care of the self-sustainability of a digitization or open access project. In this phase, the project implementing agency generates revenues from different sources such as online/print advertisements, sales promotions, copyright transfers, reprint permissions, print subscription, and consultancy services. While this phase is most dynamic, it requires a visionary leadership to formulate robust marketing strategies.

The best example of this phase is the open access initiative of Medknow Publications having interoperable collection of biomedical open access journals published from India and other countries. Medknow ensures its journal collections to be indexed in major secondary databases and metadata harvesting services. Medknow also tries to generate sustainable revenues by initiating innovative strategies in the emerging open access domain.

Model 4: Sustainability of Digitization Projects



6.3 National Information System integrating Indigenous Digital Libraries

Indian digitization and open access initiatives have definitely contributed towards development of roadmaps in achieving a national information system. The national information system will be established integrating indigenous digital libraries – developed at the national level as well as the institutional level. Over the past two decades, India is strengthening national capabilities and capacities for establishing a more inclusive national information system. Now, this national digital information system is going to be established at the zenith of an emerging knowledge economy, having characterized by achievable socio-economic and cultural indicators as elaborated in the global and national environment in Figure 6.2.

This Figure further indicates that India has an array of digitization and open access actors such as government departments, institutions of national importance, memory institutions, professional institutions, scientific societies, etc. which can play a pivotal participatory role.

This Figure also identifies several instruments for implementation of digital library and open access programmes in the country. The instruments include national/ international policy instruments, guidelines, capacity building events, and integrated curricula having components of information literacy and open source digital library software. All these instruments will definitely contribute towards the development of essential skill sets required in national intervention and implementation programmes.

This Figure also indicates a range of social benefits offered by a national digital information system that includes a vast knowledge-base supplementing to the education, training, lifelong learning and knowledge generation processes.

This Figure, derived from the present research findings, proposes an inclusive national digital information system integrating all kinds of stakeholders, policy instruments, best practices and indicators for harnessing a knowledge society.

Fig. 6.2: National Digital Information System integrating Indigenous Digital Libraries

6.4 Recommendations

This thesis tries to help in establishing a theoretical framework in the area of indigenous digital library development in a country where a wide array of documentary heritages coexists along with scholarly literature collections. A broad-based inclusive National information system integrating indigenous digital libraries is also proposed in the previous Section of this Chapter. This National information system integrates many aspects of knowledge society such as participatory development, IT localization, participatory content creation, creative content creation, and adoption of Creative Commons Licences for wide dissemination of open contents. All these are discrete elements having proposition to serve the knowledge society. Further studies can be carried out to examine the impact of their integration and convergence on the national information system.

For any advanced research, some performance indicators as well as environmental indicators are needed to be defined. These indicators can be emerged if any kind of national baseline survey is undertaken preferably by a national agency.

The sustainability of the Indian digitization initiatives is the major concern for many professionals. Long term sustainability of any information system is supplemented by innovative practices, visionary leadership and entrepreneurship capabilities of individuals as well as organizations. Thus, leadership traits and innovative practices can be measured in a comparative study that may help in elevating national capacities and capabilities.

This thesis further recommends that Indian digitization and open access projects should initiate appropriate strategic measures towards attainment of self-sustainability as depicted in the Model 4 theoretical framework.

*

ANNEX 1

Questionnaire on Content Digitization in the Library/ Information Centre/ Archive

Introduction: This questionnaire is part of a research study titled 'An evaluative study of some selected libraries in India undergoing the process of digitization' being conducted under the aegis of the Department of Library and Information Science, Jadavpur University, Kolkata. Your reply to this questionnaire is highly valuable and extremely important to assess and evaluate the state of the digitization activities in India. Information given here will be kept confidential and used for academic purposes only. Individual's identity will be kept confidential, if so desired.

ADMINISTRATIVE QUESTIONS

(I) About the Library / Institution

1.1) Name of the Library/ Information Centre/ Archive: .....

1.2) Name of the Parent Body: .....

1.3) Name of Director/ Chief Librarian/ Librarian/ In-charge: .....

1.4) Name of the Person responsible for digitization/ electronic collection: .....

1.5) Address: .....

1.6) Telephone Number: .....
Fax Number: .....
E-mail: .....

URL: ....

(II) Digitization programme and policy

2.1) Do you have a programme for digitizing collections? Yes /No

2.2) If 'Yes' go to question 4. If "No" please answer question 2.3.

2.3) Do you have a plan to undertake such a programme in near future? Yes /No

2.4) What is the timescale of the programme?.....

2.5) When did the programme begin?.....

2.6) How many items are included in the programme?.....

2.7) What are the categories of users who will use digitized documents? .....

2.8) Do you have a research programme dealing with digitization? Yes / No

2.9) If 'Yes' please provide details separately.

2.10) Are you a member of any digital library consortium(s)/ initiative(s) Yes/ No

2.11) If 'Yes' please provide details separately.

(III) Selection of materials

3.1) Which of these criteria is/are used for the selection of materials for digitization? (Tick all relevant items)

__ Materials of historical/cultural value
__ Saving of space
__ Materials of academic importance
__ Conducting research into digital processes
__ Reduction of damage
__ Preservation
__ Commercial exploitation/ use
__ Increasing visibility and access
__ Providing document delivery service
__ Other reasons (please specify) .....

(IV) Co-operation

4.1) Do you co-operate with other organisations to develop your digitization programme?
Yes / No

4.2) If yes, which kind of organizations (please tick all appropriate ones)

__ Public libraries
__ Academic libraries
__ Corporate libraries
__ Special libraries
__ Professional body's libraries
__ Archives
__ Museums
__ Corporate organizations

4.3) Is your co-operation

__ National
__ International
__ Local

4.4) Does co-operation consist of

__ Equal partnership
__ Buying services and products
__ Offering services commercially

4.5) Do you digitize material from other libraries/ archives/ information centres? Yes / No

4.6) If 'Yes', which libraries/institutions? .....

(V) Staff Pattern

5.1) Who is the person responsible for the digitization programme? .....

5.2) How many persons are engaged in the programme? (Please give full-time equivalents)
....

5.3) Please provide break ups:

Library & Information Professionals: .....
Computer and Communication Professionals: .....
Supporting Staff: .....

(VI) Costs of digitization

6.1) What is the estimated cost of digitizing documents? (please include staff, equipment, space, energy and other related items)

Per page cost (average) .....
Per book cost (average) .....
Per serial issue cost (average) .....
Cost of other items (specify) .....

6.2) How do you estimate the cost of migration/ transfer/ change to modern system? .....

TECHNICAL QUESTIONS

(VII) Digitization techniques

7.1) Is digitization carried out by	__ the Library, or	__ an outside body
7.2) Are documents prepared by	__ library staff, or	__ outside staff
7.3) Is digitization carried out from	__ the original, or	__ reproductions
7.4) If reproductions are used are they	__ photocopies __ microfilm __ slides	__ photographs __ microfiche

(VIII) Format of digitized materials

8.1) Which resolutions are used for digitization? .....

8.2) Digital image formats used __ black and white __ grey level __ colour

8.3) File format obtained	__ TIFF	__GIF	__PAL
	__ JPEG	__BMP	__ PDF
	__ others (specify).....		

8.4) Which compression methods are used. Specify.....

8.5) File size obtained after compression .....kbytes/Mbytes

8.6) Average compression rate by image types black/white..... grey.....
colour.....

8.7) Image processing software used.....

8.8) Do you digitize documents containing characters other than the Latin alphabet, especially Indian scripts? Yes / No

8.9) If 'Yes', which other scripts (please list the most important) ….....

8.10) Please give details of any special software used to digitize non-Latin scripts/ Indian scripts?
....

8.11) Do you digitize sound recordings? Yes /No
8.12) If 'Yes' what methods are used? .....
....

8.13) Do you digitize film or video? Yes / No

8.14) If 'Yes' what methods are used? .....

8.15) Have you used OCR software for texts? Yes / No

8.16) If 'Yes' which software was used?.....

8.17) Have documents undergone special treatment prior to OCR processing? Yes / No

8.18) For what purpose was OCR processing used?

__ Automatic indexing
__ Computer-assisted reading
__ Others (specify) .....

8.19) What OCR recognition rate have you obtained (by document type)? .....

8.20) Which image viewing software do you use?.....

8.21) Have you carried out any post-digitization improvements in image mode or text/OCR mode?
Yes / No

8.22) If so, which ones .....

(IX) Consultation of digitized materials

9.1) Have you developed specific workstations specially to consult the digitized collections?
Yes / No
9.2) If 'Yes' which type? .....

9.3) Which navigation software is used? .....

THE DIGITIZED DOCUMENTS

(X) The documents themselves

10.1) Are the documents:

__ isolated single documents
__ from collections

10.2) What types of documents are included in the digitization programme including percentage (%) of total documents digitized:

books (monographs)	_____%
serials	_____%
manuscripts	_____%
report literature	_____%
maps	_____%

photographs _____%
others (specify)..... _____%

10.3) Do you digitize any of the following:
__ engravings
__ prints
__ lithographs
__ posters
__ postcards
__ drawings and water-colours
__ three-dimensional objects
__ sculptures
__ fabrics and textiles
__ sound recordings
__ films and videos
__ others (specify).....

(XI) Catalogues

11.1) Are the catalogue records for digitized material
__ included in the main catalogue
__ in a separate catalogue

11.2) Is the catalogue
__ in paper form
__ electronic form
__ on an intranet server
__ available on the Internet or Websites

11.3) Are digitized materials catalogued according to a recognised standard? Yes / No

11.4) If 'Yes', which one (mention name)? .....

11.5) Are the records for digitized documents and the original
__ the same
__ independent of each other

(XII) Access to digitized documents

12.1) Are the digitized documents available
__ only on-site
__ only within the library
__ only within the institution
__ through a Website

12.2) If available through a website please give URL …….

12.3) Which functions are included in the browsing/viewing software
__ hypertext links
__ highlighting
__ others.....

12.4) Is the station equipped with
__ access control mechanisms
__ billing software

12.5) Is the workstation connected to

__ internal library servers
__ an Intranet
__ the Internet

12.6) Can users use OCR software on documents in image mode? Yes /No

12.7) What approaches for digitized documents searching are available to users?
__ Author
__ Title
__ Subject
__ Keyword
__ Series
__ Place
__ Publisher
__ Year
__ Free search
__ Boolean search

(XIII) Charges

13.1) Are users to pay to use the digitized material? Yes / No

13.2) If 'Yes', tick the appropriate items
__ on-site
__ outside the library
__ when accessed through the Website

13.3) If charges are made how are these calculated?
__ single charge
__ by time
__ by volume of material
__ by intended use (commercial/academic/students)
__ others (give details).....

13.4) If charges are made how are they collected?
__ invoice
__ cash at point of use
__ credit card
__ electronic accounting
__ others .....

(XIV) Reproduction and copyright

14.1) Do you digitize copyrighted material? Yes / No

14.2) If 'Yes', is this done
__ under legal provisions for libraries
__ with the owner's permission
__ by paying the owner a fee
__ under licence
__ without formalities

14.3) Does the library/ information centre own the copyright of the digitized form of documents?
Yes / No

14.4) If "No" who does?.....

14.5) Are users allowed to do any of the following with the digitized documents:
__ make printouts
__ download to a PC
__ download to a local network (LAN)
__ download to a general network (WAN)
__ download to a floppy/ CD-ROM

14.6) Is any electronic management system used to control copying? Yes / No

14.7) If 'Yes' which one.....

(XV) Products generated from digitized documents

15.1) Do you produce any of the following from the documents digitized
__ CD-ROM (how many?.....)
__ photographs
__ audio CDs
__paper documents

(XVI) Preservation and digitization

16.1) Do you have a preservation policy for documents in digital form? Yes / No

16.2) If a document is digitized, do you still allow access to the original? Yes / No

16.3) Are the originals stored
__ in the same way as other library materials
__ in special conditions

16.4) If in special conditions, please describe these.....

16.5) Do you have a policy for migrating data to more recent technological platforms? Yes / No

16.6) How often is data migrated?.....

16.7) Do you migrate
__ all data
__ only selected data

16.8) If "selected" how is this selection made.....
.....
.....

16.9) Is the migration automatic, based on fixed criteria (for example date) Yes / No

16.10) Is migration undertaken
__ by the library
__ an outside agency

16.11) If an outside agency, why is this?.....

(XVII) Future developments

17.1) Would you be prepared to allow your digitized documents to form part of a Digital Library?
Yes / No

17.2) Would you permit the database of such a library to be linked to your Website to permit global access to your digitized documents? Yes / No

17.3) If you would not allow this, explain why
.....
.....
.....
.....

(XVIII) Digitization programmes of other institutions in your region/ state

18.1) This questionnaire has been sent to some libraries of the institutions of national importance. Do you know of other libraries/archives in your region/ state, which have digitization programmes that would fit into this survey?
Yes / No

18.2) If 'Yes' please give their addresses
.....
.....
.....
.....
.....
.....
.....

ANNEX 2

Request for Participation (RFP) in the Digital Library of India Project

1. Name and Address of the Institution:

2. Name and address of the Coordinator:

3. Brief write up about the Institution:

4. Size of the local library:

5. List of books and other contents available locally:
(Please enclose in separate sheets)

6. Preferred Language for Indian language contents and research:

7. Number of scanners needed:

8. Number of researchers who could participate in joint research with Indian and the US side:

9. Statement of Participation:

We agree to participate in the Digital Library of India Project and agree to ensure that all the responsibilities outlined in the RFP for DLI will be fully met from our side. We also nominate ..... as the coordinator from our side.

ANNEX 3

National Mission for Manuscripts – National Survey for Manuscripts

Questionnaire

Record No.

Name of the Institute Individual Collection	
Place/ Village District State Pin Code	
Head of Institution	
Telephone Email Website	
No. of total manuscripts	
No. of paper manuscripts	
No. of palm leaf manuscripts	
No. of manuscripts in other materials	
No. of manuscripts catalogued – category-wise	
State of preservation of manuscripts	Brittle/ Good/ Bad/ Broken
Status/ Type of Institution	Govt./ Private/ Autonomous/ Others
Date on which the information is collected	

Remarks	

Subject	Language	Script	No. of Manuscripts

Signature of Surveyor

ANNEX 4

National Mission for Manuscripts

Manus Data Sheet

Record No.	
Date of Data Collection:	

Name of the Institute Personal Collection	
Village Block District State Pin Code	
Title of the Text	
Other Title	
Author	
Language	
Script	
Commentary	
Commentator	
Date of Manuscript	
Scribe	
Subject	
Bundle No.	
Manuscript No.	
No. of Folios	
Size of Manuscript	
Material	Paper/ Palm Leaf/ Birch-Bark/ Cloth/ Leather/ Other
Illustrations	
Complete/Incomplete	
Missing portion	
Condition	Brittle/ Good/ Bad/ Broken
Source of Catalogue	Descriptive/ Register/ Alphabetical/ Triennial/ Tabular/ Card Index
Remarks	

ANNEX 5

Form for Selection of Journal for Indexing in IndMED

http://indmed.nic.in

I. Journal Particulars

1. Title:

2. Former Title (if any):

3. ISSN:

4. Language:

5. Frequency:

6. Editor:

7. Postal Address for Correspondence:

8. Telephone/Fax:

9. Email:

10. Homepage (Web Site): http://

11. Year of First Publication and history of publication including non-publication:

12. Indexing Agency (if already indexed):

13. Circulation:

14. Parent body (if any):

15. Subscription Rate:

16. Whether Newsletter/In-house Publication:

II. Content Particulars

1. Subject Area of articles (content):

2. Type of Articles:

3. Brief Guidelines to authors (if any). Attach a separate sheet if required.

4. Brief description of selection procedure adopted to accept articles for publication (if any). Please also name the persons associated with the selection.

5. Please give details of publication dates of last three years and enclose the copies.

Enclosures Required:

1. Please enclose the issues of the journal mentioned above.
2. In case the journal is selected for indexing, back volumes and latest issues would be required free of cost without delay.
3. Enclose letter (on journal's letterhead) authorising NIC to index the journal.

(Signature)

Name:

Position:

ANNEX 6

List of Abbreviations

CASSIR: Cross Archive Search Services for Indian Repositories

CDAC: Centre for Development of Advanced Computing

CHDLH: Cultural Heritage Digital Library in Hindi

CIL: Cultural Informatics Lab

CoIL-Net: Content Development and IT Localization Network

CSIR: Council of Scientific and Industrial Research

DIT: Department of Information Technology

DLI: Digital Library of India

DL-RICH: Digital Library- Resource for Indian Cultural Heritage

DOAJ: Directory of Open Access Journals

DRTC: Documentation Research and Training Centre

DSIR: Department of Scientific and Industrial Research

DST: Department of Science and Technology

ETD: Electronic Theses and Dissertations

FOSS: Free and Open Source Software

GoI: Government of India

HTML: Hypertext Markup Language

IAS: Indian Academy of Sciences

ICMR: Indian Council of Medical Research

ICT: Information and Communication Technologies

IGNCA: Indira Gandhi National Centre for the Arts

IIIT: International Institute of Information Technology

IISc: Indian Institute of Sciences

IIT: Indian Institute of Technology

INSA: Indian National Science Academy

JACEP: Joint Academy's Committee for Electronic Publishing

JPEG: Joint Photographic Experts Group

MCC: Manuscript Conservation Centre

MCIT: Ministry of Communication and Information Technology

MCPC: Manuscript Conservation Partner Centre

MPC: Manuscript Partner Centre

MPEG: Moving Pictures Experts Group

MRC: Manuscript Resource Centre

NCSI: National Centre for Science Information

NDLTD: Networked Digital Library of Theses and Dissertations

NIC: National Informatics Centre

NISSAT: National Information System for Science and Technology

NMM: National Mission for Manuscripts

OAI: Open Archives Initiative

OAI-PMH: Open Archives Initiative-Protocol for Metadata Harvesting

OCR: Optical Character Recognition

OpenDOAR: Directory of Open Access Repositories

PDF: Portable Document Format

RMSC: Regional Mega Scanning Centre

TDIL: Technology Development for Indian Languages Programme

TIFF: Tagged Image File Format

TTS: Text to Speech

UNESCO: United Nations Educational, Scientific and Cultural Organization

WSIS: World Summit on the Information Society

XML: Extensible Markup Language

DR. ANUP KUMAR DAS

Websites

http://anupkumardas.blogspot.in

http://scholar.google.co.in/citations?user=Mc52nH8AAAAJ

http://in.linkedin.com/in/dranupkumardas/

http://orcid.org/0000-0001-9490-7938

www.researcherid.com/rid/F-3923-2011

http://impactstory.org/AnupKumarDas

www.researchgate.net/profile/Anup_Das7

http://jnu.academia.edu/AnupKumarDas

www.slideshare.net/anupkd